Recursive partitioning for heterogeneous causal effects

Susan Athey[a,1] and Guido Imbens[a]

[a]Stanford Graduate School of Business, Stanford University, Stanford, CA 94305

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved May 20, 2016 (received for review June 25, 2015)

In this paper we propose methods for estimating heterogeneity in causal effects in experimental and observational studies and for conducting hypothesis tests about the magnitude of differences in treatment effects across subsets of the population. We provide a data-driven approach to partition the data into subpopulations that differ in the magnitude of their treatment effects. The approach enables the construction of valid confidence intervals for treatment effects, even with many covariates relative to the sample size, and without "sparsity" assumptions. We propose an "honest" approach to estimation, whereby one sample is used to construct the partition and another to estimate treatment effects for each subpopulation. Our approach builds on regression tree methods, modified to optimize for goodness of fit in treatment effects and to account for honest estimation. Our model selection criterion anticipates that bias will be eliminated by honest estimation and also accounts for the effect of making additional splits on the variance of treatment effect estimates within each subpopulation. We address the challenge that the "ground truth" for a causal effect is not observed for any individual unit, so that standard approaches to cross-validation must be modified. Through a simulation study, we show that for our preferred method honest estimation results in nominal coverage for 90% confidence intervals, whereas coverage ranges between 74% and 84% for nonhonest approaches. Honest estimation requires estimating the model with a smaller sample size; the cost in terms of mean squared error of treatment effects for our preferred method ranges between 7–22%.

heterogeneous treatment effects | causal inference | cross-validation | supervised machine learning | potential outcomes

In this paper we study two closely related problems: first, estimating heterogeneity by covariates or features in causal effects in experimental or observational studies, and second, conducting inference about the magnitude of the differences in treatment effects across subsets of the population. Causal effects, in the Rubin causal model or potential outcome framework we use here (1–3), are comparisons between outcomes we observe and counterfactual outcomes we would have observed under a different regime or treatment. We introduce data-driven methods that select subpopulations to estimate treatment effect heterogeneity and to test hypotheses about the differences between the effects in different subpopulations. For experiments, our method allows researchers to identify heterogeneity in treatment effects that was not specified in a preanalysis plan, without concern about invalidating inference due to searching over many possible partitions.

Our approach is tailored for applications where there may be many attributes of a unit relative to the number of units observed, and where the functional form of the relationship between treatment effects and the attributes of units is not known. The supervised machine learning literature (e.g., ref. 4) has developed a variety of effective methods for a closely related problem, the problem of predicting outcomes as a function of covariates in similar environments. The most popular approaches [e.g., regression trees (5), random forests (6), LASSO (7), support vector machines (8), etc.] entail building a model of the relationship between attributes and outcomes, with a penalty parameter that penalizes model complexity. Cross-validation is often used to select the optimal level of complexity (the one that maximizes predictive power without "overfitting").

Within the prediction-based machine learning literature, regression trees differ from most other methods in that they produce a partition of the population according to covariates, whereby all units in a partition receive the same prediction. In this paper, we focus on the analogous goal of deriving a partition of the population according to treatment effect heterogeneity, building on standard regression trees (5, 6). Whether the ultimate goal in an application is to derive a partition or fully personalized treatment effect estimates depends on the setting; settings where partitions may be desirable include those where decision rules must be remembered, applied, or interpreted by human beings or computers with limited processing power or memory. Examples include treatment guidelines to be used by physicians or even online personalization applications where having a simple lookup table reduces latency for the user. We show that an attractive feature of focusing on partitions is that we can achieve nominal coverage of confidence intervals for estimated treatment effects even in settings with a modest number of observations and many covariates. Our approach has applicability even for settings such as clinical trials of drugs with only a few hundred patients, where the number of patient characteristics is potentially quite large. Our method may also be viewed as a complement to the use of "preanalysis plans" where the researcher must commit in advance to the subgroups that will be considered. It enables researchers to let the data discover relevant subgroups while preserving the validity of confidence intervals constructed on treatment effects within subgroups.

A first challenge for our goal of finding a partition and then testing hypotheses about treatment effects is that many existing machine learning methods cannot be used directly for constructing confidence intervals. This is because the methods are "adaptive": They use the training data for model selection, so that spurious correlations between covariates and outcomes affect the selected model, leading to biases that disappear only slowly as the sample size grows. In some contexts, additional assumptions such as "sparsity" (only a few covariates affect the outcomes) can be applied to guarantee consistency or asymptotic (large sample) normality of predictions (9). In this paper, we use an alternative approach that places no restrictions on model complexity, which we refer to as "honesty." We say that a model is "honest" if it does not use the same information for selecting the model structure (in our case, the partition of the covariate space) as for estimation given a model structure. We accomplish this by splitting the training sample into two parts, one for constructing the tree (including the cross-validation step) and a second for estimating treatment effects within leaves of the tree.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Drawing Causal Inference from Big Data," held March 26–27, 2015, at the National Academies of Sciences in Washington, DC. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/Big-data.

Author contributions: S.A. and G.I. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

Conflict of interest statement: The authors received funding from Microsoft Research.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: athey@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1510489113/-/DCSupplemental.

ML can also help us with treatment effect heterogeneity

ML can also help us with treatment effect heterogeneity

Use regression trees to partition units into groups with similar treatment effects

Estimation is "honest" because it uses sample splitting:

ML can also help us with treatment effect heterogeneity

Use regression trees to partition units into groups with similar treatment effects

Estimation is "honest" because it uses sample splitting:

- Split the sample in half

ML can also help us with treatment effect heterogeneity

Use regression trees to partition units into groups with similar treatment effects

Estimation is "honest" because it uses sample splitting:

- Split the sample in half
- Use one subsample to do the partitioning

ML can also help us with treatment effect heterogeneity

Use regression trees to partition units into groups with similar treatment effects

Estimation is "honest" because it uses sample splitting:

- Split the sample in half
- Use one subsample to do the partitioning
- Use the other subsample to estimate the treatment effects

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 116, NO. 536, 1716–1730: Theory and Methods Special Section on Synthetic Control Methods
https://doi.org/10.1080/01621459.2021.1891924

Matrix Completion Methods for Causal Panel Data Models

Susan Athey[a], Mohsen Bayati[b], Nikolay Doudchenko[b], Guido Imbens[c], and Khashayar Khosravi[d]

[a]Graduate School of Business, Stanford University, SIEPR, and NBER, Stanford, CA; [b]Graduate School of Business, Stanford University, Stanford, CA; [c]Graduate School of Business, and Department of Economics, Stanford University, SIEPR, and NBER, Stanford, CA; [d]Department of Electrical Engineering, Stanford University, Stanford, CA

ABSTRACT

In this article, we study methods for estimating causal effects in settings with panel data, where some units are exposed to a treatment during some periods and the goal is estimating counterfactual (untreated) outcomes for the treated unit/period combinations. We propose a class of matrix completion estimators that uses the observed elements of the matrix of control outcomes corresponding to untreated unit/periods to impute the "missing" elements of the control outcome matrix, corresponding to treated units/periods. This leads to a matrix that well-approximates the original (incomplete) matrix, but has lower complexity according to the nuclear norm for matrices. We generalize results from the matrix completion literature by allowing the patterns of missing data to have a time series dependency structure that is common in social science applications. We present novel insights concerning the connections between the matrix completion literature, the literature on interactive fixed effects models and the literatures on program evaluation under unconfoundedness and synthetic control methods. We show that all these estimators can be viewed as focusing on the same objective function. They differ solely in the way they deal with identification, in some cases solely through regularization (our proposed nuclear norm matrix completion estimator) and in other cases primarily through imposing hard restrictions (the unconfoundedness and synthetic control approaches). The proposed method outperforms unconfoundedness-based or synthetic control estimators in simulations based on real data.

ARTICLE HISTORY
Received October 2018
Accepted February 2021

KEYWORDS
Causality; Interactive fixed effects; Low-rank matrix estimation; Synthetic controls; Unconfoundedness

1. Introduction

In this article, we develop new methods for estimating average causal effects in settings with panel or longitudinal data, where some units are exposed to a binary treatment during some periods. To estimate the average causal effect of the treatment on the treated units in this setting, we impute the missing potential control outcomes.

The statistics and econometrics causal inference literatures have taken two general approaches to this problem. The literature on unconfoundedness (Rosenbaum and Rubin 1983; Imbens and Rubin 2015) can be interpreted as imputing missing potential control outcomes for treated units using observed control outcomes for control units with similar values for observed outcomes in previous periods. In contrast, the recent synthetic control literature (Abadie and Gardeazabal 2003; Abadie, Diamond, and Hainmueller 2010, 2015; Doudchenko and Imbens 2016; Chernozhukov, Wuthrich, and Zhu 2017; Ben-Michael, Feller, and Rothstein 2018; Arkhangelsky et al. 2019; Ferman and Pinto 2019; Li 2020; Amjad, Shah, and Shen 2018, see Abadie 2019 for a review) imputes missing control outcomes for treated units using weighted average outcomes for control units with the weights chosen so that the weighted lagged control outcomes match the lagged outcomes for treated units. Although at first sight similar, the two approaches are conceptually quite different in terms of the correlation patterns in the data they exploit to impute the missing potential outcomes. The unconfoundedness approach assumes that patterns over time are stable across units, and the synthetic control approach assumes that patterns across units are stable over time. In empirical work, the two sets of methods have primarily been applied in settings with different structures on the missing data or assignment mechanism. In the case of the unconfoundedness literature, the typical setting is one with the treated units all treated in the same periods, typically only the last period, and with a substantial number of control and treated units. The synthetic control literature has primarily focused on the setting with one or a small number of treated units observed prior to the treatment over a substantial number of periods. We argue that once regularization methods are used, the two approaches, unconfoundedness and synthetic controls, are applicable in the same settings, leaving the researcher with a real choice in terms of methods. In addition this insight allows for a more systematic comparison of their performance than has been appreciated in the literature.

In this study, we draw on the econometric literature on factor models and interactive fixed effects, and the computer science and statistics literatures on matrix completion, to take an approach to imputing the missing potential outcomes that is different from the unconfoundedness and synthetic control approaches. In fact, we show that it can be viewed as nesting both. In the literature on factor models and interactive effects (Bai and Ng 2002; Bai 2003) researchers model the observed outcome as the sum of a linear function of covariates and an

CONTACT Guido Imbens imbens@stanford.edu Graduate School of Business, and Department of Economics, Stanford University, SIEPR, and NBER, Stanford, CA 94305.

Causal inference is fundamentally a missing data problem

Causal inference is fundamentally a missing data problem

This is because we only ever observe $Y = D_0 Y_0 + D_1 Y_1$

Causal inference is fundamentally a missing data problem

This is because we only ever observe $Y = D_0 Y_0 + D_1 Y_1$

Athey et al. (2021) propose matrix completion methods for panel data

Causal inference is fundamentally a missing data problem

This is because we only ever observe $Y = D_0 Y_0 + D_1 Y_1$

Athey et al. (2021) propose matrix completion methods for panel data

- This is a credible data imputation technique

Causal inference is fundamentally a missing data problem

This is because we only ever observe $Y = D_0 Y_0 + D_1 Y_1$

Athey et al. (2021) propose matrix completion methods for panel data

- This is a credible data imputation technique
- Estimate the ATT by imputing Y_0 for treated units

Causal inference is fundamentally a missing data problem

This is because we only ever observe $Y = D_0 Y_0 + D_1 Y_1$

Athey et al. (2021) propose matrix completion methods for panel data

- This is a credible data imputation technique
- Estimate the ATT by imputing Y_0 for treated units
- Can be extended to take into account within-unit serial correlation