Efficient Balanced Treatment Assignments for Experimentation

SoftBlock

David Arbour[1] **Drew Dimmery**[2] Anup Rao[1]
arbour@adobe.com drewd@fb.com anuprao@adobe.com

[1] Adobe Research [2] Facebook Core Data Science

Contributions

- Reorient the problem of balancing design around a **two-sample test** between treatment and control covariate profiles.
- Motivate the **design-based** estimation of heterogeneous treatment effect.
- Show that good experimental design is governed by a **graph-cutting** problem.
- Provide an **efficient approximation** to this problem based on **maximum spanning trees**, which optimizes a ubiquitous graph-based two-sample test.

Introduction

Problem Setup

Experimental Design

- X - pre-treatment covariates
- A - binary treatment
- Y(a) - potential outcome for treatment a
- $Y = AY(1) + (1 - A)Y(0)$ - observed outcome for unit assigned treatment a

Problem Setup

Experimental Design

- X - pre-treatment covariates
- A - binary treatment
- Y(a) - potential outcome for treatment a
- $Y = AY(1) + (1 - A)Y(0)$ - observed outcome for unit assigned treatment a

Approaches:

- Simple randomization (coin flipping)
- Blocking [Greevy et al., 2004, Higgins et al., 2016]
- Rerandomization [Morgan et al., 2012, Li et al., 2018]
- Optimization [Kallus, 2018]

Problem Setup

Experimental Design

- X - pre-treatment covariates
- A - binary treatment
- Y(a) - potential outcome for treatment a
- $Y = AY(1) + (1 - A)Y(0)$ - observed outcome for unit assigned treatment a

Approaches:

- Simple randomization (coin flipping)
- Blocking [Greevy et al., 2004, Higgins et al., 2016]
- Rerandomization [Morgan et al., 2012, Li et al., 2018]
- **Optimization** [Kallus, 2018]

HTE Estimation

Estimation of HTE

The HTE for x is: $Y^{A=1}(x) - Y^{A=0}(x),\ x \in X$.

We focus on the following kernel-weighted estimator:

$$\hat{f}(x_i) = \sum_{j}^{N} \frac{\mathbb{1}(a_i \neq a_j) k(x_i, x_j) y_j}{\sum_m^N \mathbb{1}(a_i \neq a_m) k(x_i, x_m)} \tag{1}$$

which results in the following HTE estimator:

$$\hat{\tau}_i = (2a_i - 1) \left(y_i - \sum_{j=1}^{n} w_{ij} y_j \right) \tag{2}$$

Fixing observed outcomes leads to the transductive inference estimator of Zhu, Lafferty, and Ghahramani [2003].

Optimization

Defining the optimization problem for the HTE

Minimize the sum of absolute errors in each response surface:

$$\min_A \sum_i^N \left| \sum_j \frac{\mathbb{1}(a_i \neq a_j) w_{ij} y_j}{\sum_k \mathbb{1}(a_i \neq a_k) w_{ik}} - \tilde{y}_i \right| \quad (3)$$

where $\tilde{y}_i$ is the unobserved counterfactual for i.

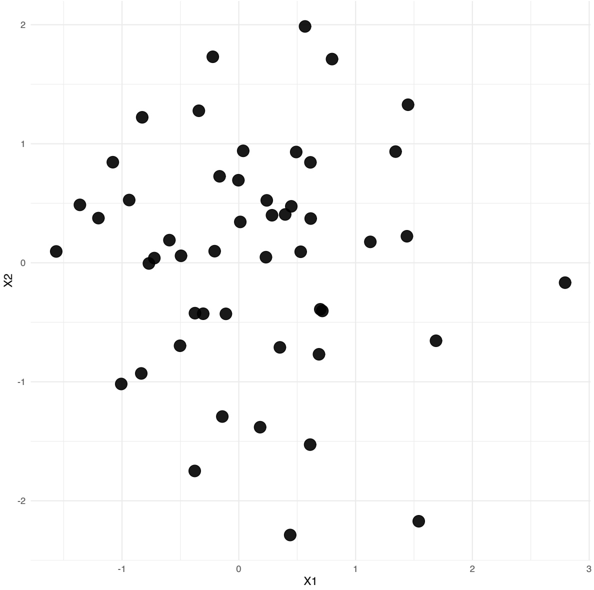
$$\sum_i^N \left| \frac{\sum_j \mathbb{1}(a_i \neq a_j) w_{ij} y_j}{\sum_k \mathbb{1}(a_i \neq a_k) w_{ik}} - y_i \right| \leq \sum_i |\tilde{y}_i - \hat{y}_i^*| + \frac{w_{sum} - \sum_{i,j} \mathbb{1}(a_i \neq a_j) w_{ij}}{d_{min}} \quad (4)$$

- w_{sum}: sum of weights
- d_{min}: minimum degree
- $\hat{y}_i^*$: weighted estimator of $\tilde{y}_i$ if all other units received treatment $1 - a_i$.

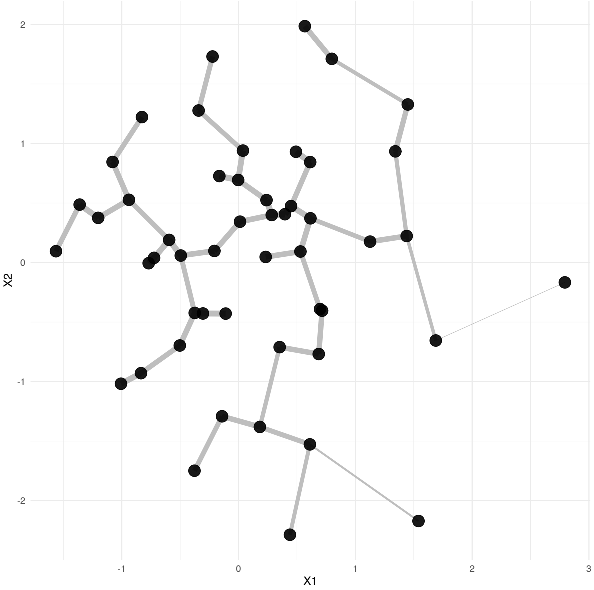
Intuition: There is irreducible error from the counterfactual, and reducible error from assignment to treatment which is minimized by maximizing $\sum_{i,j} \mathbb{1}(a_i \neq a_j) w_{ij}$.

SoftBlock

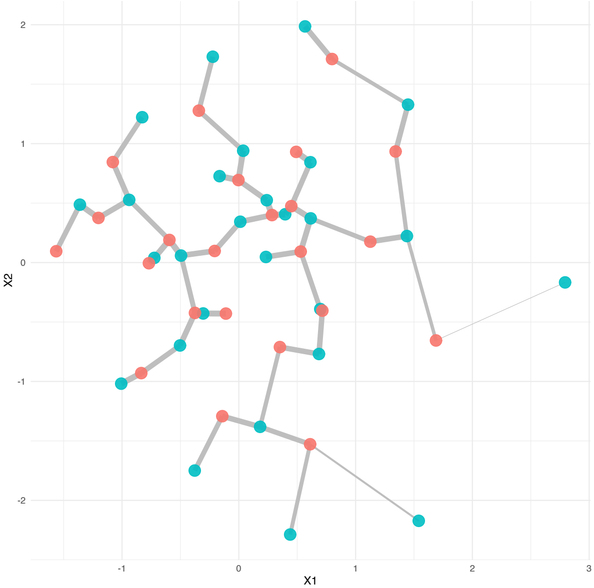
Take a point-cloud in $\mathbb{R}^2$



Find the maximum spanning tree



Greedily assign adjacent nodes alternating treatments



Algorithm

Algorithm 1: Friedman-Rafsky Minimizing Design

input : $X \in \mathbb{R}^D$

output: Assignments A, Spanning Tree T

1 $G \leftarrow$ Similarity matrix constructed from X

2 $T \leftarrow$ Maximum Spanning Tree(G)

3 $A \leftarrow$ MAXCUT(T)

- (2) is calculated in $\mathcal{O}(n \log n)$.
- (3) is calculated in $\mathcal{O}(n)$ on a tree.
- An additional algorithm (GreedyNeighbors) based on the nearest neighbor graph simply substitutes (2) for constructing the nearest neighbor graph.

Thank You!

David Arbour
arbour@adobe.com

Drew Dimmery
drewd@fb.com

Anup Rao
anuprao@adobe.com

References

Robert Greevy, Bo Lu, Jeffrey H Silber, and Paul Rosenbaum. Optimal multivariate matching before randomization. Biostatistics, 5(2):263–275, 2004.

Michael J Higgins, Fredrik Sävje, and Jasjeet S Sekhon. Improving massive experiments with threshold blocking. Proceedings of the National Academy of Sciences, 113(27): 7369–7376, 2016.

Nathan Kallus. Optimal a priori balance in the design of controlled experiments. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 80(1):85–112, 2018.

Xinran Li, Peng Ding, and Donald B Rubin. Asymptotic theory of rerandomization in treatment–control experiments. Proceedings of the National Academy of Sciences, 115 (37):9157–9162, 2018.

Kari Lock Morgan, Donald B Rubin, et al. Rerandomization to improve covariate balance in experiments. The Annals of Statistics, 40(2):1263–1282, 2012.

Xiaojin Zhu, John Lafferty, and Zoubin Ghahramani. Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In ICML 2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining, volume 3, 2003.