Max Planck Institute for Intelligent Systems – Empirical Inference

Conditional Distributional Treatment Effect with Kernel Conditional Mean Embeddings and U-Statistic Regression

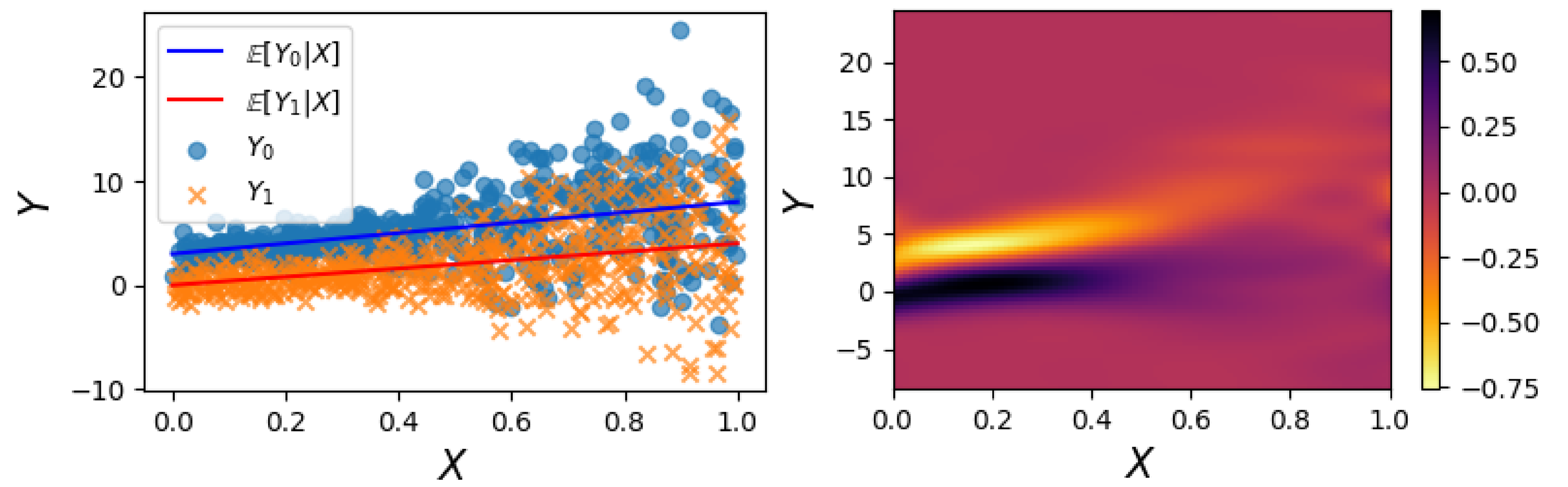
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

Junhyung Park, Uri Shalit, Bernhard Schölkopf and Krikamol Muandet

Introduction & Motivation

Treatment effect analysis has long been a central topic in a wide range of domains, including econometrics, political sciences, healthcare and social sciences.

ATE and CATE, which stand for *average treatment effect* and *conditional average treatment effect* respectively, have long been used to quantify the effect of a treatment. However, they fail to capture *distributional aspects* of the treatment effect.



As we can see in the first plot, comparison of conditional means is only meaningful if the corresponding variances are taken into account. This is one compelling reason to look beyond the mean, but other aspects of the distributions can also be investigated.

Summary of contributions:

- Formal definition of the *conditional distributional treatment effect* (CoDiTE), associated with a chosen distance function between distributions.
- Test of equality between the control and treatment distributions conditioned on the covariates, using *conditional mean embeddings*.
- Exploratory analysis of where and by how much the control and treatment densities differ, conditioned on the covariates, using *conditional witness functions*.
- Treatment effect analysis with respect to specific distributional quantities, such as conditional variance, using *u-statistic regression*, an extension of kernel ridge regression.

Potential Outcomes: Notations & Assumptions

Notations

- $(\Omega, \mathcal{F}, P)$ is the underlying probability space.
- $\mathcal{X}$ is the input space and $\mathcal{Y} \subseteq \mathbb{R}$ is the output space.
- $Z : \Omega \to \{0, 1\}$ represents treatment assignment (e.g. drug assignment).
- $X : \Omega \to X$ represents the covariates (e.g. patient characteristics).
- $Y_0, Y_1 : \Omega \to \mathcal{Y}$ represent potential outcomes under control and treatment assignment, respectively (e.g. blood pressure).
- $Y = Y_0(1 - Z) + Y_1 Z$ represents the observed outcome, i.e. the potential outcome consistent with the treatment assignment.

Assumptions

- **Unconfoundedness:** Z is independent of (Y_0, Y_1) given X.
- **Overlap:** $0 < e(X) = P(Z = 1 \mid X) = \mathbb{E}[Z \mid X] < 1$.

Core Definitions

Conditional Distributional Treatment Effect (CoDiTE)

Let D be some distance function between probability measures. We define the *conditional distributional treatment effect* (CoDiTE) associated with D as

$$U_D(x) = D\left(P_{Y_0|X=x}, P_{Y_1|X=x}\right).$$

The choice of D here should depend on what aspects of distribution is desired to be captured.

Precision of Estimating Heterogeneous Distributional Effects (PEHDE)

Given a distance function D, for an estimator $\hat{U}_D$ of U_D, we define the *precision of estimating heterogeneous distributional effects* (PEHDE) as

$$\psi_D\left(\hat{U}_D\right) = \left\|\hat{U}_D - U_D\right\|_2^2 = \mathbb{E}\left[\left|\hat{U}_D(X) - U_D(X)\right|^2\right].$$

The two questions we answer in this paper are as follows:

Q1: Are $P_{Y_0|X=x}$ and $P_{Y_1|X=x}$ different, i.e. is there *any* distributional effect of the treatment?
Q2: If so, *how* does the distribution of the treatment group differ from the control group?

CoDiTE associated with MMD via CMEs

In order to answer **Q1**, we choose D to be the *maximum mean discrepancy* (MMD):

$$U_{\text{MMD}}(x) = \text{MMD}\left(P_{Y_0|X=x}, P_{Y_1|X=x}\right) = \left\|\mu_{Y_1|X=x} - \mu_{Y_0|X=x}\right\|_{\mathcal{H}},$$

where $\mu_{Y_0|X=x}$ and $\mu_{Y_1|X=x}$ are the *conditional mean embeddings* of Y_0 and Y_1 given X respectively. If the MMD is associated with a characteristic kernel, $P_{Y_0|X=x}$ and $P_{Y_1|X=x}$ are equal if and only if $U_{\text{MMD}}(x) = 0$.

Hypotheses

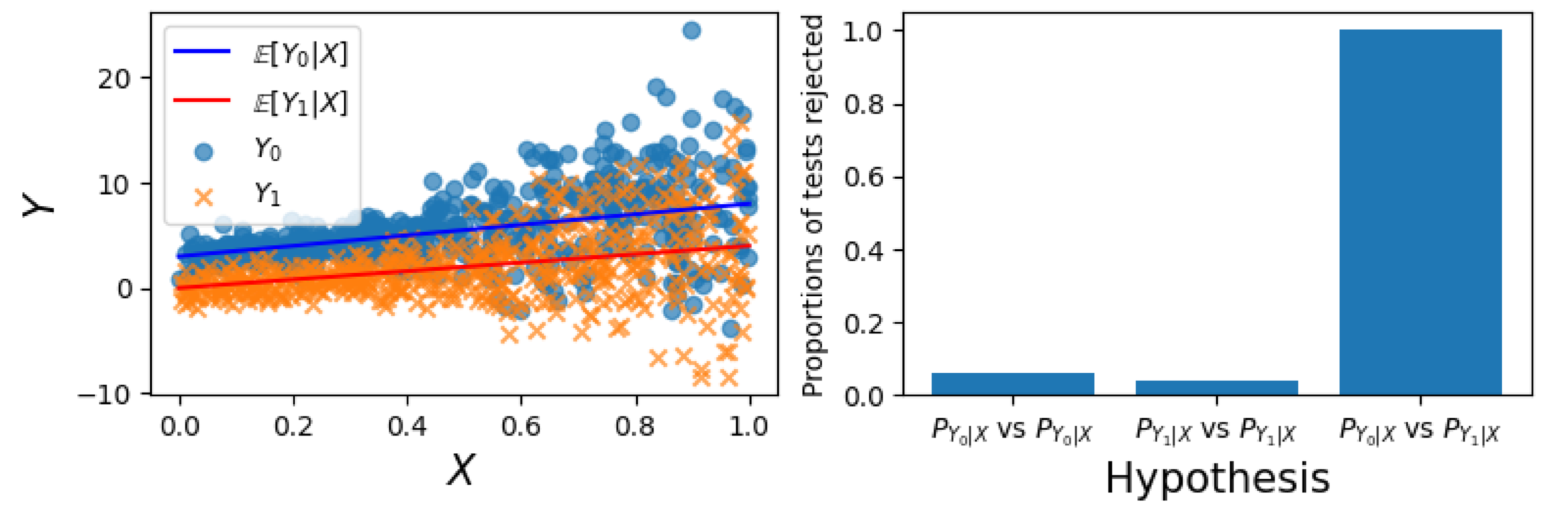
Consider the following null and alternative hypotheses:

- H_0: $P_{Y_0|X=x} = P_{Y_1|X=x}$ P_X-almost everywhere.
- H_1: There exists $A \subseteq \mathcal{X}$ with positive measure such that $P_{Y_0|X=x} \neq P_{Y_1|X=x}$ for all $x \in A$.

We can integrate $U_{\text{MMD}}(x)$ over $\mathcal{X}$ to obtain a test statistic for these hypotheses:

$$t = \mathbb{E}\left[\left\|\mu_{Y_1|X} - \mu_{Y_0|X}\right\|_{\mathcal{H}}^2\right].$$

Then, using a conditional resampling scheme with this test statistic, we can perform a statistical hypothesis test to test for these hypotheses.

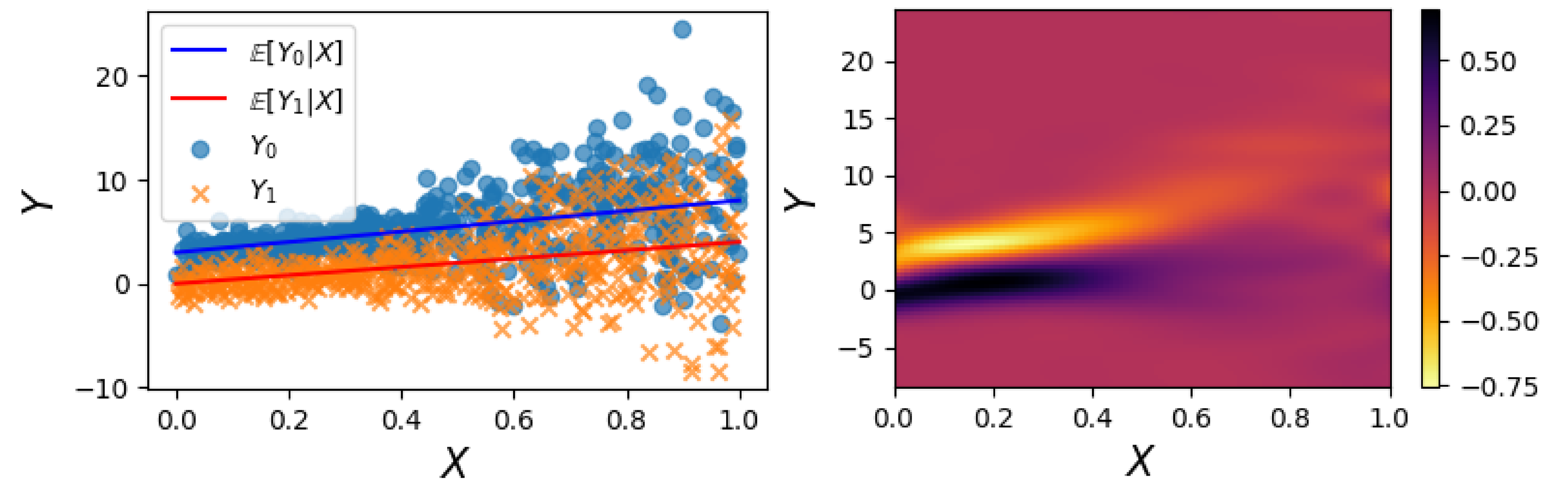


Conditional Witness Functions

The **conditional witness function** between $P_{Y_1|X=x}$ and $P_{Y_0|X=x}$ is

$$\mu_{Y_1|X=x} - \mu_{Y_0|X=x} : \mathcal{Y} \to \mathbb{R}.$$

- For $y \in \mathcal{Y}$ in regions where the density of $P_{Y_1|X=x}$ is greater than that of $P_{Y_0|X=x}$ we have $\mu_{Y_1|X=x}(y) - \mu_{Y_0|X=x}(y) > 0$.
- For y in regions where the converse is true, we similarly have $\mu_{Y_1|X=x}(y) - \mu_{Y_0|X=x}(y) < 0$.
- The greater the difference in density, the greater the magnitude of the witness function.



Kernel U-Statistic Regression

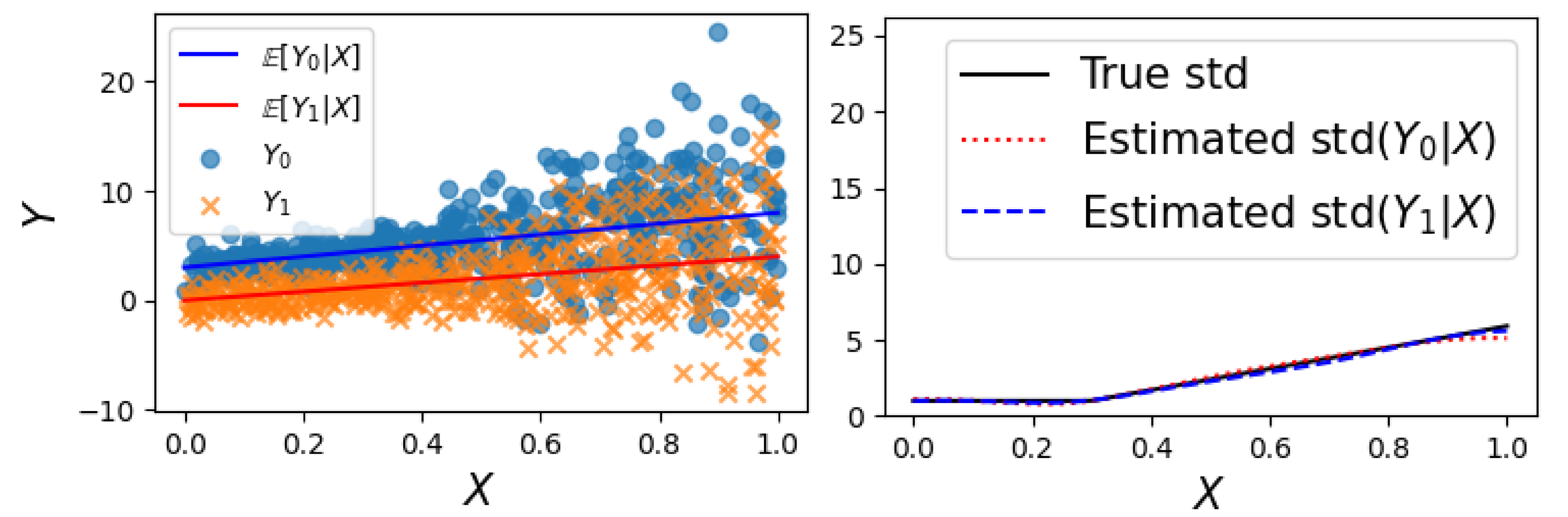
Many interesting properties of a distribution, such as mean, variance and skewness, can be represented as a *U-statistic*. Let $Y_1, ..., Y_r$ be independent copies of Y, and let $h : \mathcal{Y}^r \to \mathbb{R}$ be a symmetric funciton. U-statistics are used to estimate functions of the form

$$\mathbb{E}\left[h\left(Y_1, ..., Y_r\right)\right].$$

We propose *kernel U-statistic regression* to estimate the conditional counterparts, for each of the control and treatment groups, such as the conditional variance:

$$\theta\left(P_{Y_0|X}\right) = \mathbb{E}\left[h\left(Y_{01}, ...Y_{0r}\right) \mid X_1, ..., X_r\right], \qquad \theta\left(P_{Y_1|X}\right) = \mathbb{E}\left[h\left(Y_{11}, ...Y_{1r}\right) \mid X_1, ..., X_r\right]$$

This is done by extending the usual kernel ridge regression for estimating the conditional mean. Regression is performed in $\mathcal{H}^r$, the r-times tensor product of the original RKHS.



Conditional Distributional Treatment Effect with Kernel Conditional Mean Embeddings and U-Statistic Regression
Junhyung Park, Uri Shalit, Bernhard Schölkopf and Krikamol Muandet
Paper ID: 873
Link to paper