Robust and Agnostic Learning of Conditional Distributional Treatment Effects

Nathan Kallus, Miruna Oprescu

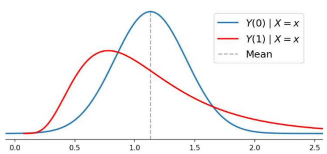
Paper Code

Setup

- Binary treatment $A \in \{0,1\}$, covariates $X \in \mathcal{X}$, potential outcomes $Y(0), Y(1) \in \mathbb{R}$
- Data: $Z_i = (X_i, A_i, Y_i) \sim Z = (X, A, Y(A))$
- Ignorability: $Y(a) \perp\!\!\!\perp A \mid X$
- Overlap: $e^*(X) = \mathbb{P}(A=1 \mid X)$ $e^*(X) \in (0,1)$

Motivation

- Skewed outcome functions (e.g., income, platform usage) and risk quantification



Potential outcomes with the same conditional mean but different tail effects.

- Need to look beyond the conditional mean effect: Conditional Distributional Treatment Effects (**CDTEs**)

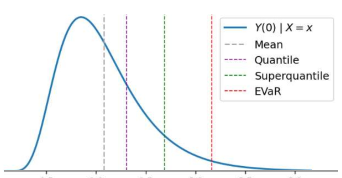
CDTEs

$$\mathrm{CDTE}(X) = \kappa^*(F_{Y(1)|X}) - \kappa^*(F_{Y(0)|X})$$

where $\kappa^*(F)$ is any distribution statistic.

Example of statistics and corresponding CDTEs:

- Mean (CATE)
- Quantile (CQTE)
- Superquantile, i.e., tail average (CSQTE)
- *f*-risk measures based on *f*-divergences, e.g. entropic-value-at-risk, or EVaR (C*f*RTE)



Different distribution statistics (quantile, superquantile, EVaR) at level 0.75.

CDTE Estimation Algorithm

General Framework: Moment Statistics

- Consider statistics that solve moment equations:

$$\mathbb{E}_F[\rho(Y, \kappa, h)] = 0$$

where $h^*(F)$ is a set of nuisances.

Plugin Estimator

$$\mathrm{CDTE}^{\mathrm{Plugin}}(X) = \widehat{\kappa}_1(X) - \widehat{\kappa}_0(X)$$

where $\widehat{\kappa}_a(\cdot)$ are estimates for $\kappa_a(\cdot)$.

- Inherits bias from the nuisances $\widehat{\kappa}_a(\cdot)$.
- Can wash out the signal when the nuisances are more complex than the CDTE.

Pseudo-outcome Regression Estimator

1. Estimate the nuisances $\nu_a^* = (\kappa_a^*, h_a^*)$.
2. Get pseudo-outcome derived from the efficient influence function (EIF):

$$\psi(Z, e, \alpha, \nu) = \kappa_1(X) - \kappa_0(X) - \frac{A - e(X)}{e(X)(1-e(X))}\alpha_A(X)^T \rho(Y, \nu_A(X))$$

where $\alpha_A(X)$ are additional nuisances to estimate.

3. Regress pseudo-outcome on covariates $X \in \mathcal{X}$.

Algorithm 1 CDTE Learner

Input: Data $\{(X_i, A_i, Y_i) : i \in \overline{1,n}\}$, folds $K \geq 2$, nuisance estimators, regression learner

```
for k ∈ 1,K do
    Use data {(X_i, A_i, Y_i) : i ≠ k − 1 (mod K)} to
    construct nuisance estimates ê^(k), α̂^(k), ν̂^(k)
    for i = k − 1 (mod K) do
        Set ψ̂_i = ψ(Z_i, ê^(k), α̂^(k), ν̂^(k))
    end for
end for
return CDTE^(x) = Ê_n[ψ̂ | X = x]
```

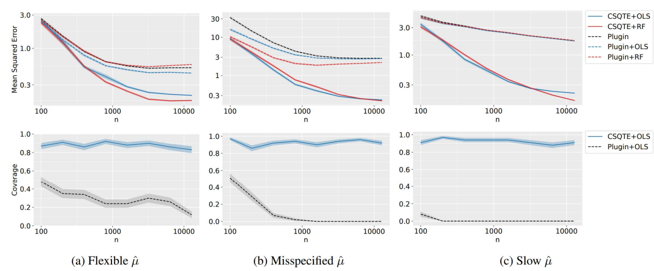
Learning and Inference Guarantees

Robustness:

- The error has a product structure so small errors in the nuisances lead to second-order errors in the CDTE estimates.
- E.g., if nuisances are estimated at a rate of at least $O(n^{-1/4})$, CDTEs are estimated at the rate $O(n^{-1/2})$
- There are many chances at consistency when some of the nuisances are misspecified.

Model Agnostic:

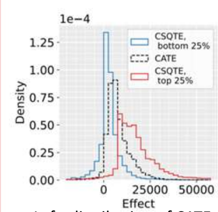
- Linear regression parameters are asymptotically normal with oracle variance (i.e., if we use OLS as the final stage, the confidence intervals are valid)



Performance of CSQTE learner with flexible, misspecified or slow converging $\kappa_a(\cdot)$ estimator. Either a flexible last stage (random forest) or OLS is used.

Case Study: Effect of 401k Eligibility

- Effect of 401k eligibility on net worth
- CSQTE on bottom and top 25% asset holders



Coefficient	CSQTE Bottom 25%	CATE	CSQTE Top 25%
Intercept ($10,000)	−0.021 (−1.06, 1.02)	−0.95 (−2.42, 0.51)	−2.07 (−7.04, 2.90)
Income	0.25** (0.08, 0.43)	0.21 (−0.08, 0.50)	−0.05 (−1.12, 1.01)
Age	105 (−75, 286)	232** (24, 441)	513 (−182, 1210)
Education	−801** (−1440, −164)	16 (−1050, 1090)	1340 (−2490, 5180)

Left: distribution of CATEs and CSQTEs with random forest last stage. Right: linear regression coefficients with OLS final stage.