Jakob R. Juergens
University of Wisconsin – Madison
Department of Economics

Inference for Conditional Average Treatment Effects using Distributional Nearest Neighbors

Key Contributions

- Nonparametric second stage for the DML2-Estimator using Generalized U-Statistics
- Improved Bias-Variance Tradeoffs for DNN Estimator

Nonparametric Regression Setup

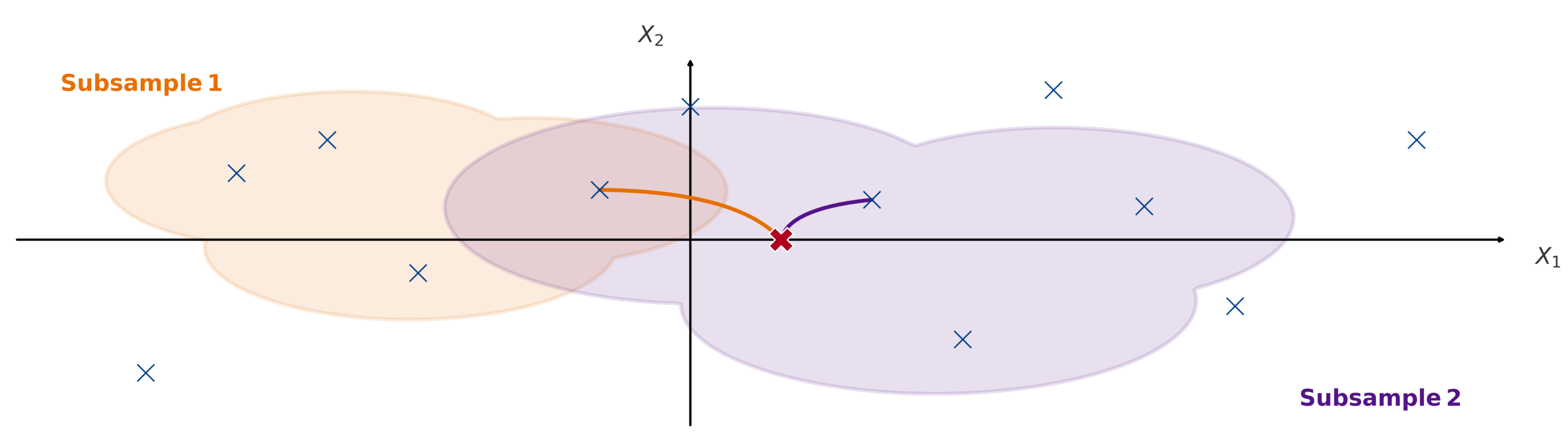
The observed data consists of an i.i.d. sample $\mathbf{D}_n = \{Z_i = (X_i, Y_i)\}_{i=1}^n$ from

$$Y = \mu(X) + \varepsilon \quad \text{with} \quad \mathbb{E}[\varepsilon \mid X] = 0, \quad \mathsf{Var}(\varepsilon \mid X = x) = \sigma_\varepsilon^2(x)$$

DNN-Regression Estimator

To estimate the Regression function at a point of interest x:

1. Order the sample based on the distance to the point of interest.
$$||X_{(1)} - x|| \leq ||X_{(2)} - x|| \leq \ldots \leq ||X_{(n)} - x||$$
Let $\mathsf{rk}(x; X_i, D)$ denote this *rank* assigned to observation i in a sample D
2. Define a data-driven kernel function
$$K(x, X_i) = \binom{n}{s}^{-1} \sum_{\ell \in L_{n,s}} 1\left(\mathsf{rk}(x; X_i, \mathbf{D}_\ell) = 1\right)$$
where $L_{n,s} = \{(l_1, \ldots, l_s) \in [n] \mid l_1 < l_2 < \ldots < l_s\}$.
3. Construct the estimator $\widehat{\mu}_s(x)$ as the solution to
$$0 = \sum_{i=1}^{n} K(x, X_i)\left(Y_i - \widehat{\mu}(\mathbf{x})\right) = \sum_{i=1}^{n-s+1} \frac{\binom{n-i}{s-1}}{\binom{n}{s}} \left(Y_{(i)} - \widehat{\mu}_s(\mathbf{x})\right)$$



Asymptotic Normality and Jackknife Consistency

Under some mild assumptions, for any fixed x, we have for some positive sequence ω_n of order $\sqrt{s/n}$

$$\left(\widehat{\mu}_s(x) - \mu(x)\right)/\omega_n \rightsquigarrow \mathcal{N}(0, 1)$$

as $n, s \to \infty$ with $s = o(n)$. Furthermore, the Jackknife variance estimator is ratio consistent:

$$\left(\frac{n-1}{n} \sum_{i=1}^{n} \left(\widehat{\mu}_s(x; \mathbf{D}_{n,-i}) - \widehat{\mu}_s(x; \mathbf{D}_n)\right)^2\right) \Big/ \omega_n^2(x) \longrightarrow_p 1.$$

CATE Estimation Setup

The observed data consists of an i.i.d. sample $\mathbf{D}_n = \{Z_i = (X_i, W_i, Y_i)\}_{i=1}^n$ from

$$Y = 1(W = 0)\mu_0(X) + 1(W = 1)\mu_1(X) + \varepsilon, \quad W_i \sim \mathrm{Bern}(\pi_0(X_i))$$
$$\text{with} \quad \varepsilon \perp\!\!\!\perp W \mid X, \quad \mathbb{E}[\varepsilon \mid X] = 0, \quad \mathsf{Var}(\varepsilon \mid X = x) = \sigma_\varepsilon^2(x)$$

DNN-DML2 CATE Estimator

To estimate the Conditional Average Treatment Effect at a point of interest x:

1. Take an equally sized K-fold partition $\mathcal{I} = (I_k)_{k=1}^K$ of $[n]$. For each $k \in [K]$, define $I_k^C = [n] \backslash I_k$. For observation j denote the corresponding fold by k_j. For the observation being assigned rank i, denote the corresponding fold by $k_{(i)}$.
2. For each $k \in [K]$, use a first-stage estimator on the data set $\mathbf{D}_{I_k^C}$ to estimate the nuisance parameters
$$\widehat{\mu}_k^w(x) = \widehat{\mu}\left(x; \mathbf{D}_{I_k^C}^{(w)}\right) \quad \text{for } w = 0, 1, \quad \widehat{\pi}_k(x) = \widehat{\mu}\left(x; \mathbf{D}_{I_k^C}\right) \quad \text{where the predicted variable is } W$$
3. Construct the estimator $\widehat{\theta}(x)$ as the solution to
$$0 = \sum_{k=1}^{K} \sum_{i \in I_k} K(x, X_i) m\left(Z_i; \widehat{\theta}(x), \widehat{\eta}_k\right) = \sum_{i=1}^{n-s+1} \frac{\binom{n-i}{s-1}}{\binom{n}{s}} m\left(Z_{(i)}; \widehat{\theta}(x), \widehat{\eta}_{k_{(i)}}\right)$$
where
$$m(Z; \theta(x), \eta) = \mu_1(X) - \mu_0(X) + \left(\frac{W}{\pi(X)} - \frac{1 - W}{1 - \pi(X)}\right) \times (Y - \mu_W(X)) - \theta(X)$$

DDML Rate-Assumptions

Let $\{\delta_n\}_{n \geq 1}$ and $\{\Delta_n\}_{n \geq 1}$ be some of positive sequences converging to zero such that $\delta_n \geq n^{-1/2}$. Given a random $I \subset [n]$ with $|I| = n/k$, we have $P\left(\widehat{\eta}(\mathbf{D}_{I^C}) \in \mathcal{T}_n\right) \geq 1 - \Delta_n$, where $\eta_0 \in \mathcal{T}_n$ and

$$r_n' := \sup_{\eta \in \mathcal{T}_n} \left(\mathbb{E}_Z\left[\left|m(Z; \theta_0, \eta) - m(Z; \theta_0, \eta_0)\right|^2\right]\right)^{1/2} \leq \delta_n,$$
$$\lambda_n' := \sup_{r \in (0,1), \eta \in \mathcal{T}_n} \left|\partial_r^2 \mathbb{E}_Z\left[m(Z; \theta_0, \eta_0 + r(\eta - \eta_0))\right]\right| \leq \delta_n / \sqrt{n}.$$

Asymptotic Normality

Under some mild smoothness assumptions on the potential outcome functions and the propensity score, for any fixed x, we have for some positive sequence ω_n of order $\sqrt{s/n}$

$$\left(\widehat{\theta}(x) - \theta_0(x)\right)/\omega_n \rightsquigarrow \mathcal{N}(0, 1)$$

as $n, s \to \infty$ with $s = o\left(\min\left\{1/r_n', \left(\sqrt{n} \cdot \lambda_n'\right)^{-1}\right\}\right)$ with rates corresponding to the above [DDML-Rate Assumptions].