Is It Correlation or Causation?
Uncertainty Quantification in Estimating Causal Effects with Unobserved Confounding

Miruna Oprescu* (2nd Year Fellow)

Introduction: Causal Effects

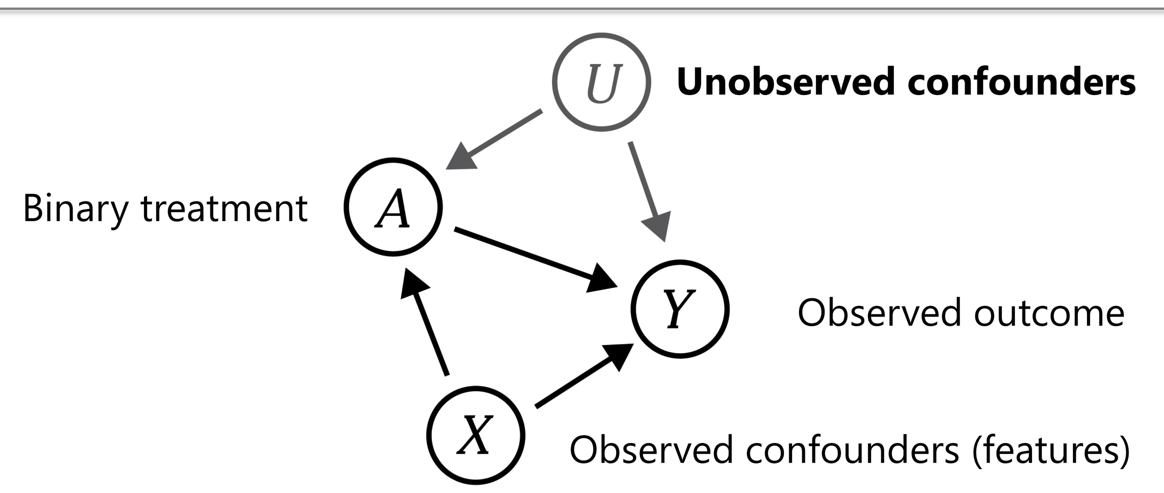
- Setup: binary treatment $A \in \{0,1\}$, features $X \in \mathcal{X}$, potential outcomes $Y(0), Y(1) \in \mathbb{R}$ under A.
- We only observe data $Z_i = (X_i, A_i, Y_i) \sim (X, A, Y(A))$.
- Conditional average treatment effect (CATE):

$$\tau(x) = \mathbb{E}[Y(1) \mid X = x] - \mathbb{E}[Y(0) \mid X = x]$$

- Example: effect of ibuprofen on headaches.

Unit: X	Treatment: A	Pain Score: Y	Pain Score: Y
		6	4
		7	3
		9	7
		5	1

Setup: Unobserved Confounding



- If $U \neq \emptyset$, we can't tell between causation and correlation!
- The best we can do: find **bounds** $\tau^+(x), \tau^-(x)$ on $\tau(x)$.

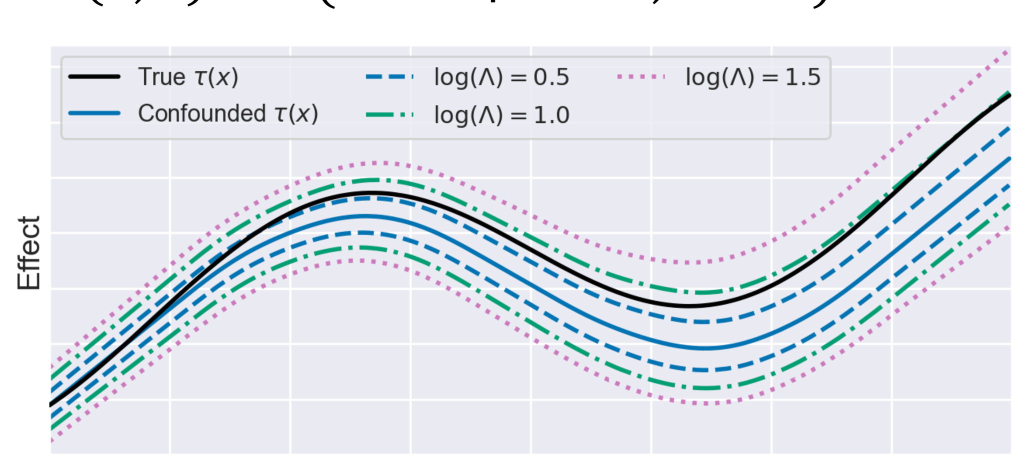
Uncertainty Quantification

Assumption: Marginal Sensitivity Model (MSM).

$$\Lambda^{-1} \leq \frac{e(x,u)}{1-e(x,u)} \Big/ \frac{e(x)}{1-e(x)} \leq \Lambda$$

where $e(x) = P(A = 1 \mid X = x)$

$e(x,u) = P(A = 1 \mid X = x, U = u)$



Example of CATE bounds under different values of Λ.

- Cornfield et al. (1959) found that confounding would need to be 9 times larger in smokers than non-smokers to negate the observed effect of smoking on lung cancer (Λ=9).

TL;DR

We propose the **B-Learner**, a flexible meta-learner that learns **bounds on causal effects** under unobserved confounding. The B-Learner's bounds are **valid** (correct with high probability), **sharp** (tightest possible), and **efficient** (requires less data).

$\tau^\pm(x)$ Bounds With The Marginal Sensitivity Model

Let:

$\mu(x,a) = \mathbb{E}[Y \mid X = x, A = a]$

$Y^\pm(x,a) = \sup/\inf \mathbb{E}[Y(a) \mid X = x]$

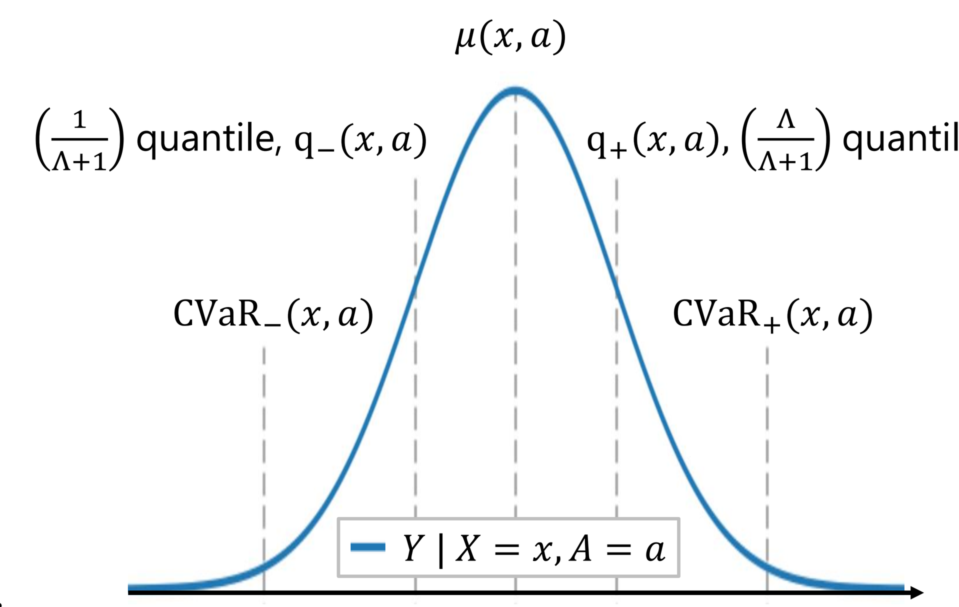
Then:

$$\boldsymbol{Y^+(x,1)} = e(x)\mu(x,1) + (1-e(x))\rho_+(x,1)$$

$$\boldsymbol{Y^-(x,0)} = (1-e(x))\mu(x,0) + e(x)\rho_-(x,0)$$

$$\boldsymbol{\tau^+(x)} = \boldsymbol{Y^+(x,1)} - \boldsymbol{Y^-(x,0)}$$

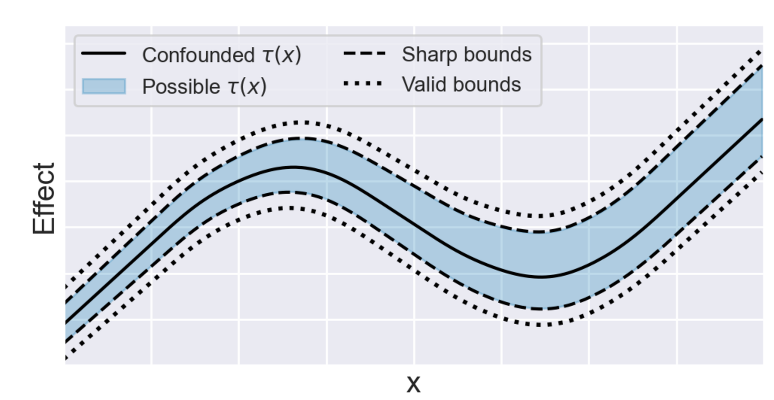
s.t. $\rho_\pm(x,a) = \Lambda^{-1}\mu(x,a) + (1-\Lambda^{-1})CVaR_\pm(x,a)$.



B-Learner: Efficient Estimation of CATE Bounds

Bound **estimates** should be...

- **Valid:** correct with high probability.
- **Sharp:** tightest possible.
- **Efficient and Robust:** Converge with as little data as possible regardless of models used.



Attempt #1: Naïve "Plug-in" Estimator

Estimate $e(x), \mu(x,a), \rho_\pm(x,a)$ and "plug" them into $Y^\pm(x,a)$ to obtain:

$$\hat{\tau}^+_{\text{Plugin}}(x) = \hat{Y}^+(x,1) - \hat{Y}^-(x,0)$$

- Inherits bias from the estimated nuisances $\hat{e}(x), \hat{\mu}(x,a), \hat{\rho}_\pm(x,a)$ which means that it cannot guarantee the desired bound properties.

Attempt #2: The B-Learner Algorithm

1. Estimate nuisance set $\hat{\eta} = (\hat{e}(x), \hat{q}_\pm(x,a), \hat{\rho}_\pm(x,a))$ in one sample.
2. Derive a debiasing term for the plug-in estimator via the efficient influence function:

$$\phi^+_\tau(Z,\hat{\eta}) = \underbrace{\hat{\tau}^+_{\text{Plugin}}(X)}_{\text{"plug-in"}} - \underbrace{f(\hat{\eta}(X,A))}_{\text{bias correction}}$$

   where $f(\cdot)$ is a known function (that is too complex to write out).

3. Regress pseudo-outcome $\phi^+_\tau(Z,\hat{\eta})$ on features $X \in \mathcal{X}$ in another sample.

Algorithm 1 The B-Learner

input Data $\{(X_i, A_i, Y_i) : i \in \{1, ..., n\}\}$, folds $K \geq 2$, nuisance estimators, regression learner $\widehat{\mathbb{E}}_n$

for $k \in \{1, ..., K\}$ **do**

  Use data $\{(X_i, A_i, Y_i) : i \neq k-1 \pmod K\}$ to construct nuisance estimates $\widehat{\eta}^{(k)} = (\widehat{e}^{(k)}, \widehat{q}^{(k)}, \widehat{\rho}^{(k)})$

  for $i = k-1 \pmod K$ **do**

    Set $\widehat{\phi}^+_{\tau,i} = \phi^+_\tau(Z_i, \widehat{\eta}^{(k)})$

  end for

end for

output $\widehat{\tau}^+(x) = \widehat{\mathbb{E}}_n[\widehat{\phi}^+_\tau \mid X = x]$

Theoretical Guarantees

(Machine Learning jargon)

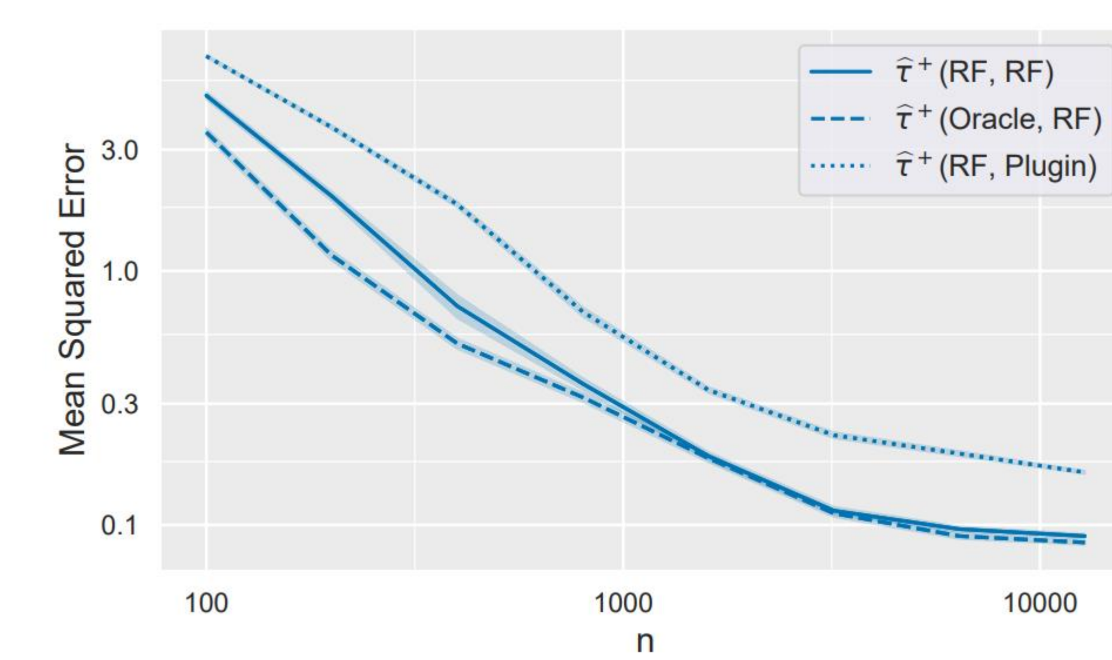
- $\mathbf{L_2}$ validity, sharpness and robustness guarantees for **ERM** second stage estimator $\widehat{\mathbb{E}}_n$:
  1. L_2 bias on the order of
     $$\mathcal{E} = \|\hat{e} - e\|_{L_2}\|\hat{\rho} - \rho\|_{L_2} + \|\hat{q} - q\|^2_{L_2}.$$
  2. If $\hat{q}$ and either $\hat{e}$ or $\hat{\rho}$ are consistent, the bounds are **sharp** on average.
  3. If $\hat{q}$ is inconsistent, the bounds are still **valid** in expectation.
  4. The B-Learner has **quasi-oracle efficiency,** i.e. it learns bounds at a statistical rate influenced by the complexity of the target class.
- **Pointwise** validity, sharpness and robustness guarantees for **linear smoother** second stage estimator $\widehat{\mathbb{E}}_n$.

Experiments

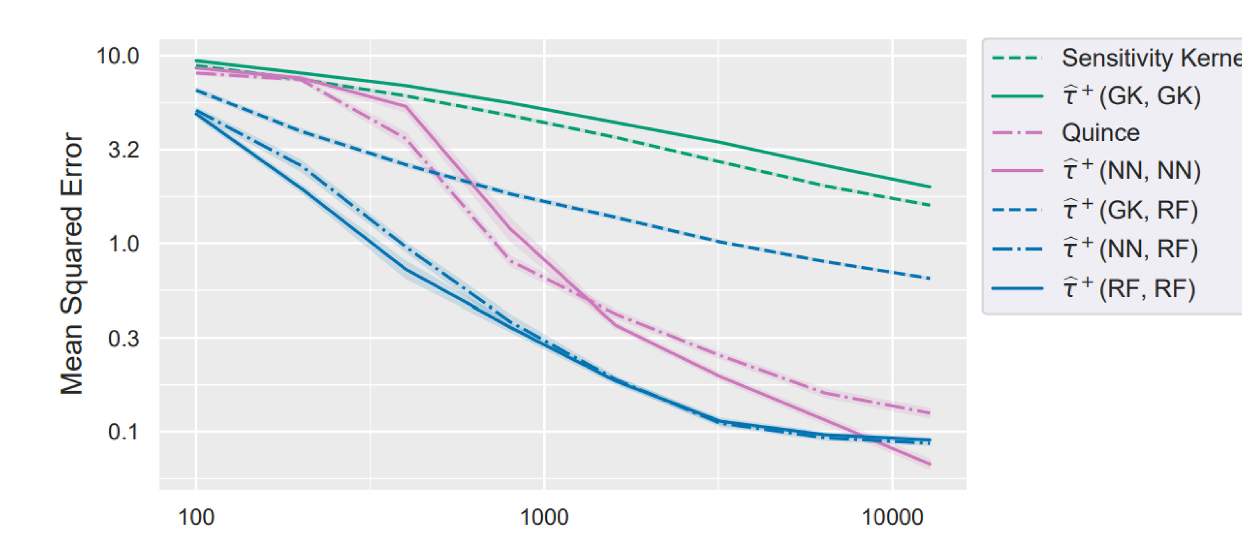
Simulations

$$A \sim \text{Bernoulli}(\text{logit}(0.75X_0 + 0.5))$$

$$Y \sim \mathcal{N}((2A-1)(X_0+1) - 2\sin((4A-2)X_0), 1)$$



Quasi-oracle property of the B-Learner algorithm. n is the sample size. In $\hat{\tau}^+(x,y)$, x and y are the types of first- and second-stage nuisances.



Performance of the B-Learner compared with the *Sensitivity Kernel* (Kallus et al. 2019) and *Quince* (Jesson et al., 2021). GK=Gaussain Kernel, NN=Neural Network, RF=Random Forest.

* Joint work with J. Dorn, M. Ghoummaid, A. Jesson, N. Kallus, U. Shalit. "B-learner: Quasi-oracle bounds on heterogeneous causal effects under hidden confounding." International Conference on Machine Learning. PMLR, 2023.