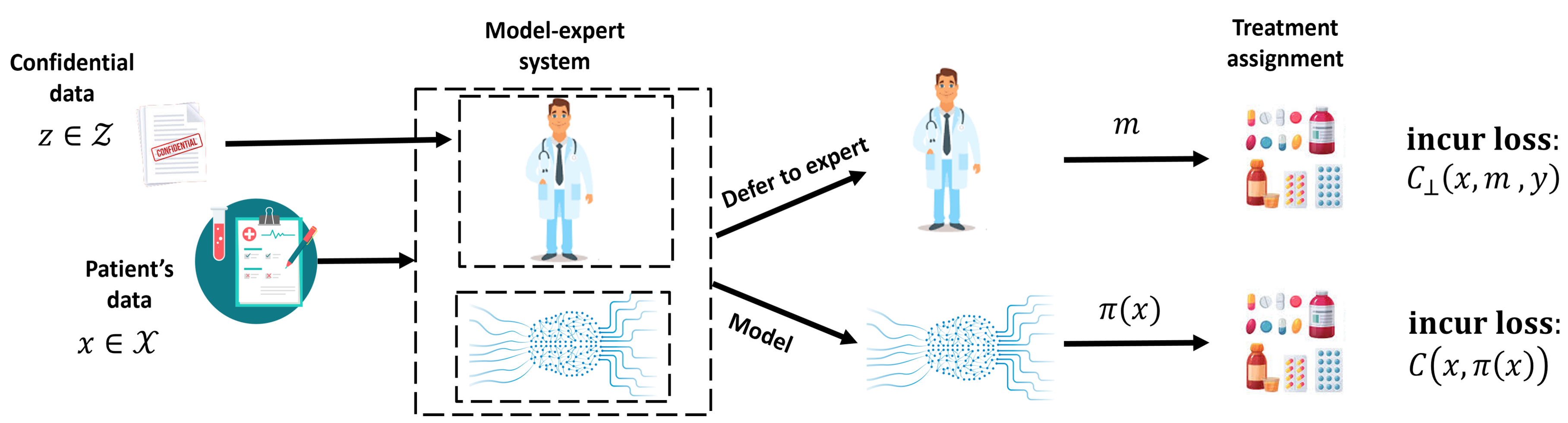
When to Act and When to Ask: Policy Learning With Deferral Under Hidden Confounding

Marah Ghoummaid, Uri Shalit

The Faculty of Data and Decision Sciences, Technion

Goal

Learning a policy with deferral for treatment recommendation from observational data under hidden confounding.



Problem Setup

- Observational data under the Neyman-Rubin potential outcomes framework [Rubin, 2005].
- **Data Distribution:**
$$(X, A, Y(1), Y(0), U) \sim P_{full}$$
- **Observed Data**: $(X, A, Y) \sim P$, with $Y = Y(A)$.
- **Task:** Learn a **policy with deferral** $\pi: \mathcal{X} \to \{0,1,\perp\}$, where $\perp$ means deferral to an expert.

Conditional Average Potential Outcomes (CAPOs)

Given $X = x$, and a treatment $A = a$, CAPO is defined as:
$$Y(x, a) = \mathbb{E}[Y(a)|X = x]$$

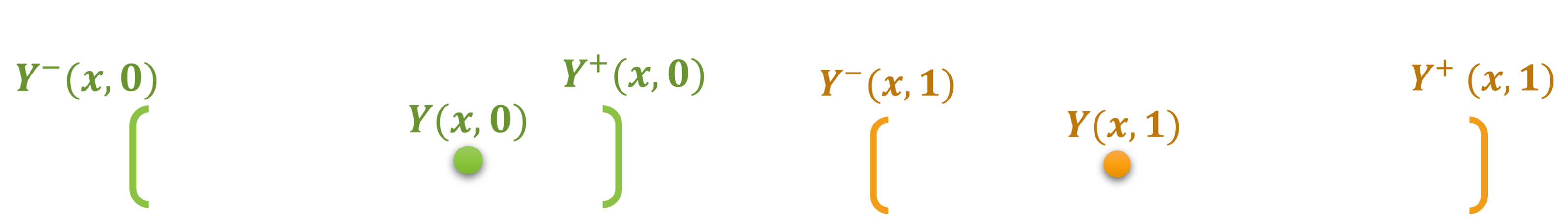
Marginal Sensitivity Model (MSM)

Assumption: There exists $\Lambda \geq 1$ such that the following holds almost surely under P_{full}:
$$\Lambda^{-1} \leq \frac{e(x,u)}{1-e(x,u)} \Big/ \frac{e(x)}{1-e(x)} \leq \Lambda$$

- $e(x) = P(A = 1|X = x)$ - the observed propensity score,
- $e(x, u) = P_{full}(A = 1|X = x, U = u)$ - the full propensity score, where U is the hidden confounder.

CAPO Bounds

Let $\mathcal{M}(\Lambda)$ the set of distributions consistent with the observed data (X,A,Y) and the MSM, then:
$$Y^+(x, a) = \max_{Q \in \mathcal{M}(\Lambda)} \mathbb{E}[Y(a)|X = x]$$
$$Y^-(x, a) = \min_{Q \in \mathcal{M}(\Lambda)} \mathbb{E}[Y(a)|X = x]$$



CAPO-Based Policies

Bounds Policy
$$\pi^{\hat{Q}}_{\text{bounds}}(x) = \begin{cases} 1 & \text{if } \hat{Y}^-(x,1) - \hat{Y}^+(x,0) > 0 \\ 0 & \text{if } \hat{Y}^+(x,1) - \hat{Y}^-(x,0) < 0 \\ \perp & \text{otherwise} \end{cases}$$

Pessimistic Policy
$$\pi^{\hat{Q}}_{\text{pessimistic}}(x) = \begin{cases} 1 & \text{if } \hat{Y}^-(x,1) - \hat{Y}^+(x,0) > 0 \\ 0 & \text{if } \hat{Y}^+(x,1) - \hat{Y}^-(x,0) < 0 \\ 1 & \text{otherwise, if } \hat{Y}^-(x,1) - \hat{Y}^-(x,0) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Cost-sensitive Objective

$$L(\pi) = \mathbb{E}_{(x,y)\sim P, m\sim M|(x,y)}\left[C(x, \pi(x))\mathbb{I}_{\pi(x)\neq\perp} + C_\perp(x, m, y)\mathbb{I}_{\pi(x)=\perp}\right]$$

Challenge:

$L(\pi)$ is non-convex and computationally hard to optimize

Surrogate Loss Function

(Building on Mozannar and Sontag, 2020)

Policy: $\pi_i: \mathcal{X} \to \mathbb{R}$ where $\pi(x) = \underset{i\in\{0,1,\perp\}}{\operatorname{argmin}}\, \pi_i(x)$.

CAPO Bounds: $\hat{Q}(x) = \left(\hat{Y}^+(x,0), \hat{Y}^-(x,0), \hat{Y}^+(x,1), \hat{Y}^-(x,1)\right)$

Costs: $c(0) = C(x, 0)$, $c(1) = C(x, 1)$, and $c(\perp) = C_\perp(x, m, y)$.

Weights: $w^j(z, \hat{Q}(x)) = \max_{k\in\{0,1,\perp\}} c(k) - c(j)$.

Surrogate loss function for L:
$$L_{CE}(\pi, z; \hat{Q}) = \sum_{j\in\{0,1,\perp\}} -w^j\left(z, \hat{Q}(x)\right) \log\left(\frac{\exp(\pi_j(x))}{\sum_{k\in\{0,1,\perp\}} \exp(\pi_k(x))}\right)$$

Conservative Costs

$$C(x, 1) = Y^+(x, 0) - Y^-(x, 1)$$
$$C(x, 0) = Y^+(x, 1) - Y^-(x, 0)$$
$$C_\perp(x_i, a, y_i) = \begin{cases} Y^-(x,0) - y & , if\ a = 1 \\ Y^-(x,1) - y & , otherwise \end{cases}$$

Theoretical Guarantees

CAPO Bounds Estimation: using the B-Learner (Oprescu et al., 2023), with guarantees on validity and convergence rates.

Proven Properties (under mild assumptions on policy learners):

Cor 1. (Consistency): the surrogate loss L_{CE} achieves the same optimum as the machine-expert loss L.

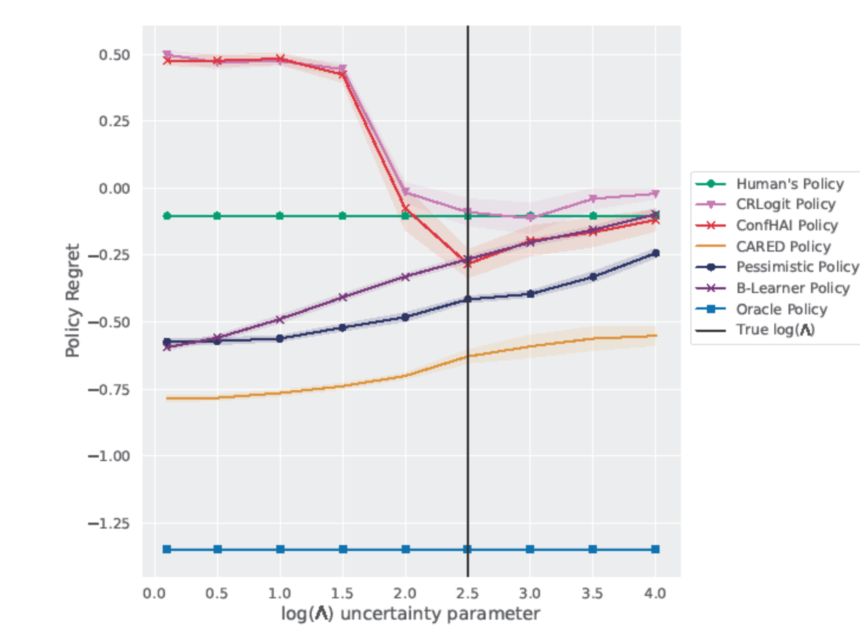
Thm 1. (Costs Are Coherent): minimizing costs in L_{CE} ensures decisions are non-inferior to those by the expert or machine alone.

Thm 2. (Generalization Bound): a generalization bound is provided for L_{CE} .

Experiment

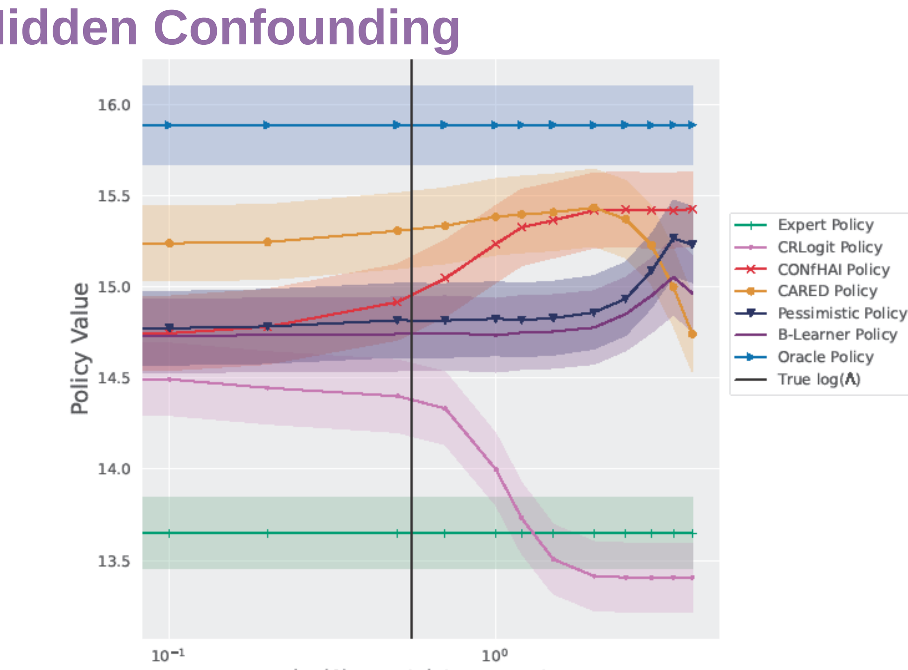
Synthetic Data

$$\xi \sim Bern(0.5), X \sim \mathcal{N}((2\xi - 1)\mu_x, I_5),$$
$$U = \mathbb{I}[Y(1) < Y(0)],$$
$$Y(A) = \beta_0^T x + \mathbb{I}[A = 1]\beta_{treat}^T x + 0.5\alpha\xi\mathbb{I}[A = 1] + \eta + \omega\xi + \epsilon$$
$$\beta_0 = [0, 0.5, -0.5, 0, 0], \beta_{treat} = [-1.5, 1, -1, -1.5, 1, 0.5],$$
$$\mu_x = [-1, 0.5, -1, 0, -1], \eta = 2.5$$
$$\alpha = -2, \omega = 1.5, \text{ and } \epsilon \sim \mathcal{N}(0,1).$$
$$e(x) = \sigma(\beta^T X) \text{ with } \beta = [0.075, -0.5, 0, -1, 0].$$
$$e(X, U) = \frac{(\Lambda_0 U + 1 - U)e(X)}{[1 + 2(\Lambda_0 - 1)e(X) - \Lambda_0]U + \Lambda_0 + (1 - \Lambda_0)e(X)}$$, with the true Λ_0 such that $\log(\Lambda_0) = 2.5$.

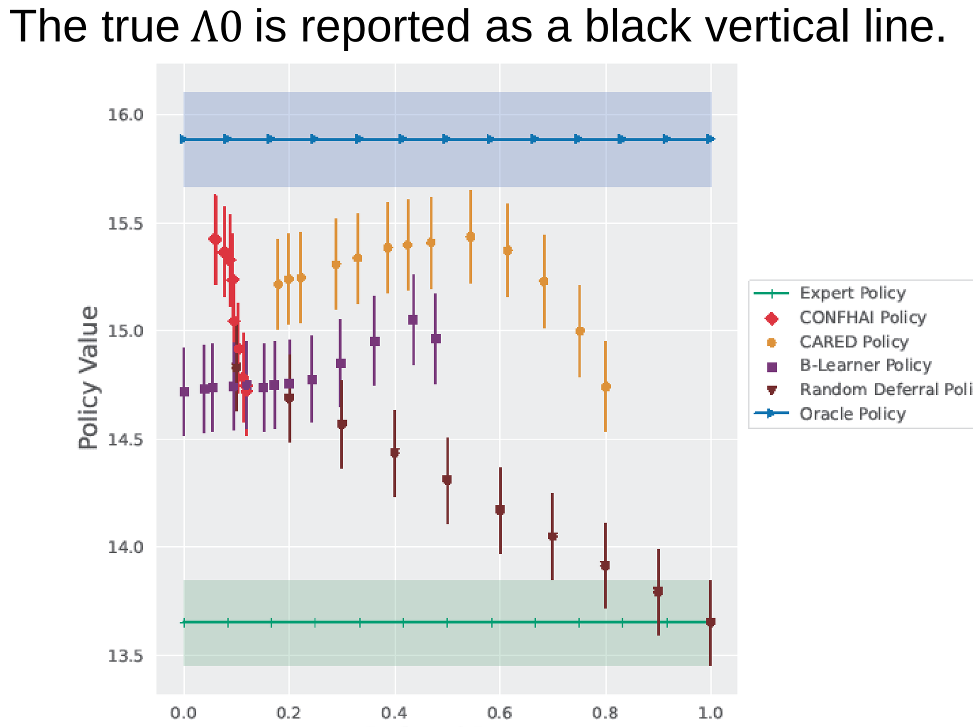


Policy regret for different levels of hidden confounding (MSM). lower policy regret is better. The true Λ0 is reported as a black vertical line. **Human's Policy:** the human expert's (A) in the observed data, **CRLogit Policy:** [Kallus and Zhou, 2020] **ConfHAI Policy:** [Gao and Yin, 2023], **CARED(ours), Pessimistic Policy** and **B-Learner Policy:** CAPO-based from the B-Learner [Oprescu et al., 2023],. **Oracle Policy:** the best true policy.

IHDP Hidden Confounding



Policy regret for different levels of hidden confounding (MSM). The true Λ0 is reported as a black vertical line.



Policy value for different rates of deferral.
Random Deferral Policy: that defers a randomly chosen fraction of samples to the expert at each deferral rate.