# Efficient Policy Learning from Surrogate-Loss Classifications

Andrew Bennett (Cornell Tech)

Joint work with Nathan Kallus (Cornell Tech)

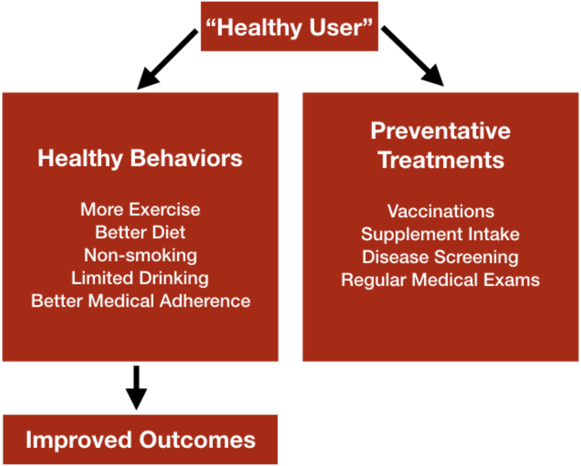
# This Talk

# Offline Policy Learning Problem


| Child Weight | Infant Drops (50 mg / 1.25 ml) | Children's Suspension (100 mg / 5 ml) | Jr. Strength Tablets (100 mg) | Adult Tablets |
|---|---|---|---|---|
| 12 – 17 lbs. (6 mos. & older) | 1.25 ml | ½ teaspoon (2.5 ml) | — | — |
| 18 – 23 lbs. | 1.875 ml | ¾ teaspoon (3.75 ml) | — | — |
| 24 – 35 lbs. | 2.5 ml | 1 teaspoon (5 ml) | 1 tablet | — |
| 36 – 47 lbs. | — | 1 ½ teaspoon (7.5 ml) | 1 ½ tablets | — |
| 48 – 59 lbs. | — | 2 teaspoon (10 ml) | 2 tablets | 1 tablet |

**Important problem since experimenting with treatments may be unethical, costly, or impossible!**

# Offline Policy Learning Problem




**Problem is non-trivial since data may be confounded!**

# Our Contribution

- Although there is work on efficient policy evaluation, we show that performing policy optimization using an efficiently estimated objective **does not** lead to efficient learning of optimal policy parameters.
- We present a novel algorithm for efficiently learning optimal policy parameters, and show that it leads to improved regret.

# This Talk

# Policy Learning Formalization

- Notation:
  - $X$: Individual covariates
  - $T$: Binary treatment decision ($-1$ or 1)
  - $Y$: Observed outcome given actual treatment $T$
  - $Y(t)$: **(unobserved)** Potential outcome that would have occurred under treatment $t$, for $t \in \{-1, 1\}$
- Policy: $\pi(x)$ denotes treatment decision given $X = x$.
- Assume iid data $((X_1, T_1, Y_1), \dots, (X_n, T_n, Y_n))$.
- **Goal:** Find a policy $\pi$ that maximizes $\mathbb{E}[Y(\pi(X))]$.
- Assume that $Y = Y(T)$, and that logging policy was a function of $X$ only (no hidden confounding)

# Policy Learning Objective

- Let $J(\pi) = \mathbb{E}[Y(\pi(X))] - \frac{1}{2}[Y(+1) - Y(-1)]$
- Can easily show that $J(\pi) = \mathbb{E}[\psi\pi(X)]$ for a range of $\psi$ such as
  - $\psi_{\text{IPS}} = \frac{TY}{P(T|X)}$
  - $\psi_{\text{DM}} = \mathbb{E}[Y \mid X, T = 1] - \mathbb{E}[Y \mid X, T = -1]$
  - $\psi_{\text{DR}} = \psi_{DM} + \psi_{IPS} - \frac{T\mathbb{E}[Y|X,T]}{P(T|X)}$
- Given estimates of the nuisance functions $P(T = t \mid X = x)$ and $\mathbb{E}[Y \mid X = x, T = t]$, can estimate $\psi$ from the observed data for any given triplet $(X, T, Y)$
- This suggests approximating the above objective by $J_n(\pi) = \frac{1}{n}\sum_{i=1}^{n} \hat{\psi}_i \pi(X_i)$, where $\hat{\psi}_i$ are estimated as above.
- **Reduction to Empirical Risk Minimization!**

# Surrogate-Loss Reduction

- $J_n(\pi) = \frac{1}{n}\sum_{i=1}^{n} \hat{\psi}_i \pi(X_i)$ is a non-convex objective.
- As in (weighted) binary classification, we can make the problem tractable by replacing this $0/1$ loss with a convex surrogate.
- Consider parametric class of functions $g_\theta$ for $\theta \in \Theta$, and let $\pi_\theta(x) = \text{sign}(g_\theta(x))$ denote parametric policy class.
- Surrogate-loss objective is

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} |\hat{\psi}_i| l(g_\theta(X_i), \text{sign}(\hat{\psi}_i)) \,,$$

for some convex loss function $l$

# Efficient Policy Learning

- There is a rich past literature that considers efficiently estimating the cost function to be optimized, based on the above approaches.
- **However, none of this work addresses whether the optimal policy parameters themselves are efficiently estimated!**

# This Talk

# Assumptions for Surrogate-Loss Reduction

1. **Valid Weighting Function:**

$$\mathbb{E}[\psi \mid X] = \mathbb{E}[Y(+1) - Y(-1) \mid X]$$

   - This holds for the IS, DM, and DR $\psi$ functions.

2. **Regularity:**

$$\mathbb{E}[|\psi|] < \infty$$

3. **Correct Specification:**

$$\{g_\theta : \theta \in \Theta\} \cap \left( \underset{g \text{ unconstrained}}{\arg\min} \mathbb{E}[|\psi| l(g(X), \text{sign}(\psi))] \right) \neq \varnothing$$

   - This assumption means there is a policy in our parametric class that minimizes the population surrogate loss.

# Semiparametric Model Implied by Assumptions

**Lemma**

*Assuming correct specification, and using logistic regression loss, the implied model is given by all distributions on $(X, T, Y)$ for which there exists $\theta^* \in \Theta$ satisfying*

$$\frac{P(\psi > 0 \mid X)}{\mathbb{E}[|\psi| \mid X]} = \sigma(g_{\theta^*}(X)) \quad \textit{almost surely.}$$

- $\sigma$ denotes the logistic function.
- This model is in general *semiparametric*, because the set of distributions on $(X, T, Y)$ that satisfy this constraint is infinite-dimensional.

# Efficient Learning Implies Optimal Regret

- Recall that $J$ is true objective, and $L$ is surrogate loss objective. Define:

$$\text{Regret}_J(\theta) = \underset{\pi \text{ unconstrained}}{\arg\max} \; J(\pi) - J(\theta)$$
$$\text{Regret}_L(\theta) = L(\theta) - \inf_{\theta \in \Theta} L(\theta)\,,$$

- and let $AR_L(\hat{\theta}_n)$ be the distribution limit of $n\text{Regret}_L(\hat{\theta}_n)$.

**Theorem**

*Under our assumptions, we have $\text{Regret}_J(\theta) \leq \text{Regret}_L(\theta)$. Furthermore, if $\hat{\theta}_n$ is a semiparametrically efficient then $\mathbb{E}[\phi(AR_L(\hat{\theta}_n))]$ is minimized for any non-decreasing $\phi$.*

# This Talk

# Conditional Moment Formulation

**Theorem**

*Define*

$$m(X;\theta) = \mathbb{E}[|\psi| l'(g_\theta(X), \mathrm{sign}(\psi)) \mid X] \,.$$

*Then under our assumptions the policy given by* $\theta^*$ *is optimal if and only if* $m(X;\theta^*) = 0$ *almost surely.*

- This is equivalent to $\mathbb{E}[f(X)|\psi| l'(g_\theta(g_\theta(X), \mathrm{sign}(\psi))] = 0$ for every square integrable function $f$.
- There exists an extensive literature on solving these kinds of problems efficiently!

# ESPRM Algorithm

- We extend an existing algorithm that was previously designed for efficiently solving instrumental variable regression.
- Define:

$$u(X, \psi; \theta, f) = f(X)|\psi| l'(g_\theta(X), \text{sign}(\psi))$$

$$U_n(\theta, f; \tilde{\theta}) = \frac{1}{n} \sum_{i=1}^{n} u(X_i, \hat{\psi}_i; \theta, f) - \frac{1}{4n} \sum_{i=1}^{n} u(X_i, \hat{\psi}_i; \tilde{\theta}, f)^2$$

- Given some flexible function class $\mathcal{F}$, and prior policy estimate $\tilde{\theta}$, our *efficient surrogate policy risk minimization* (ESPRM) estimator is given by:

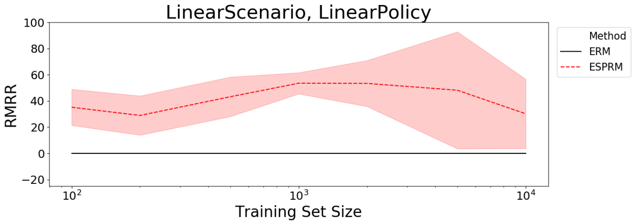
$$\hat{\theta}^{\text{ESPRM}} = \underset{\theta \in \Theta}{\arg\min} \sup_{f \in \mathcal{F}} U_n(\theta, f; \tilde{\theta})$$

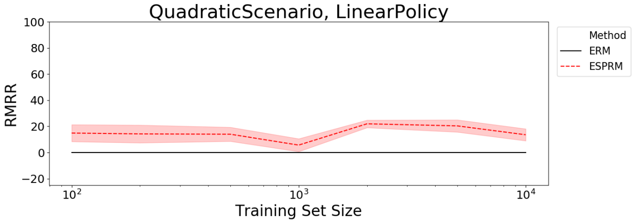
# This Talk

# Experimental Setup

- We compare our ESPRM method against empirical risk minimization (ERM) using the logistic regression surrogate loss.
- We test the following policy classes:
  - LinearPolicy: $g_\theta(x) = \theta^t x$
  - FlexiblePolicy: $g_\theta$ parametrized by neural network
- We test on the following kinds of synthetic scenarios:
  - Linear: Optimal policy is linear
  - Quadratic: Optimal policy is quadratic
- In all cases we compare algorithms over a large number of randomly generated synthetic scenarios of the given kind.

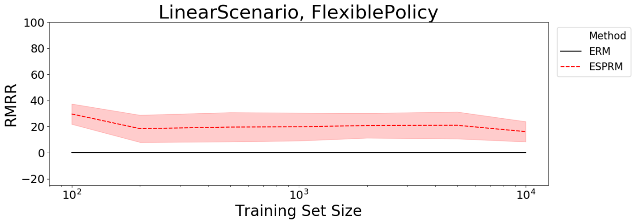
# Experimental Results – Linear Policy Class

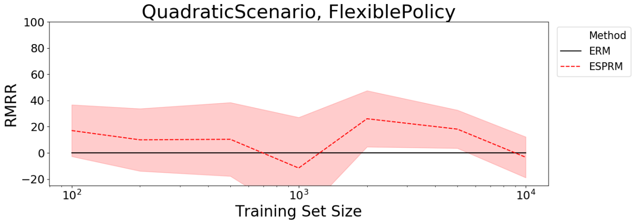



$$RMRR(\hat{\theta}_n) = \left(1 - \frac{\mathbb{E}[\text{Regret}_J(\hat{\theta}_n)]}{\mathbb{E}[\text{Regret}_J(\hat{\theta}_n^{\text{ERM}})]}\right) \times 100\%,$$

# Experimental Results – Flexible Policy Class




$$RMRR(\hat{\theta}_n) = \left(1 - \frac{\mathbb{E}[\text{Regret}_J(\hat{\theta}_n)]}{\mathbb{E}[\text{Regret}_J(\hat{\theta}_n^{\text{ERM}})]}\right) \times 100\%,$$

# Jobs Case Study

- We consider a case study based on different programs assigned to unemployed individuals in France.
- Individuals were randomly assigned to either an intensive counseling program by a private agency, or a similar program by a public agency.
- We also have access to individual covariates and outcomes (based on whether they re-entered work force within six months, minus treatment cost).

# Jobs Case Study

- We divide the experimental data into train and test splits, and introduce selection bias by randomly dropping training units based on covariates.
- We learn treatment assignment policies using ESPRM and ERM on the artificially confounded training data.
- Learnt policies are evaluated on the test data using a Horvitz-Thompson estimator.
- Estimated policy values:

| **Policy Class** | **ERM** | **ESPRM** | **Difference** |
|---|---|---|---|
| Linear | $-0.96 \pm 4.32$ | $4.42 \pm 3.78$ | $5.38 \pm 5.06$ |
| Flexible | $-1.75 \pm 4.64$ | $7.68 \pm 3.16$ | $9.42 \pm 5.17$ |

# Summary

- Although there is work on efficient policy evaluation, policy learning using an efficiently estimated objective **does not** lead to efficient learning of optimal policy parameters.
- We presented an algorithm for policy learning based on theory of conditional moment problems that is efficient.
- We showed both theoretically and empirically that efficient optimal policy estimation implies improved regret.

# Thank You

Thank you for listening, and please check our our full paper "Efficient Policy Learning from Surrogate-Loss Classification Reductions"!