Policy Estimation

Susan Athey
Stanford University

See Athey and Wager (2017)

Efficient Policy Estimation

- **Learning optimal policy** assignment and **estimating treatment effect heterogeneity** closely related
- ML literature proposed variety of methods (Langford et al; Swaminathan and Joachims; in econometrics, Kitagawa and Tetenov, Hirano and Porter, Manski)
- Estimating the value of a personalized policy closely related to **estimating average treatment effect** (comparing treat all policy to treat none policy)
- Lots of econometric theory about how to estimate average treatment effects efficiently (achieve semi-parametric efficiency bound)
- Athey and Wager (2017): prove that bounds on regret (gap between optimal policy and estimated policy) can be tightened using an algorithm consistent with econometric theory
  - Theory provides guidance for algorithm choice

Setup and Approach

Setup

- Policy $\pi : \mathcal{X} \to \{\pm 1\}$
- Given a class of policies Π, the optimal policy π^* and the regret $R(\pi)$ of any other policy are respectively defined as

$$\pi^* = \operatorname{argmax}_{\pi \in \Pi} \left\{ \mathbb{E}\left[Y_i\left(\pi\left(X_i \right) \right) \right] \right\} \tag{1}$$

$$R(\pi) = \mathbb{E}\left[Y_i\left(\pi^*\left(X_i \right) \right) \right] - \mathbb{E}\left[Y_i\left(\pi\left(X_i \right) \right) \right]. \tag{2}$$

- Goal: estimate a policy π that minimizes regret $R(\pi)$.

Approach: Estimate $Q(\pi)$, choose policy to minimize $\hat{Q}(\pi)$:

$$Q(\pi) = \mathbb{E}\left[Y_i\left(\pi\left(X_i \right) \right) \right] - \frac{1}{2} \mathbb{E}\left[Y_i(-1) + Y_i(+1) \right] \tag{3}$$

$$\hat{\pi} = \operatorname{argmax}_{\pi \in \Pi} \left\{ \widehat{Q}(\pi) \right\}, \tag{4}$$

Alternative Approaches

$$Q(\pi) = \mathbb{E}\left[Y_i\left(\pi\left(X_i\right)\right)\right] - \frac{1}{2}\mathbb{E}\left[Y_i(-1) + Y_i(+1)\right] \tag{5}$$

Methods for estimating ATE or ATT can also be used to estimate the effect of any policy–simply define treatment as following policy, and control as the reverse

Kitagawa and Tetenov (forthcoming, Econometrica):

- Estimate $\hat{Q}(\pi)$ using inverse propensity weighting
- Suppose that $Y_i \leq M$ uniformly bounded, overlap satisfied with $\eta \leq e(x) \leq 1 - \eta$, Π is Vapnik-Chervonenkis Class of dimension VC(Π)
  - Roughly, VC dimension is the maximum integer D such that some data set of cardinality D can be "shattered" by a policy in Π. If tree is limited to D leaves, its VC dimension is D.
- Then, $\hat{\pi}$ satisfies the regret bound

$$R\left(\hat{\pi}\right) = \mathcal{O}_P\left(\frac{M}{\eta}\sqrt{\frac{\mathrm{VC}\left(\Pi\right)}{n}}\right). \tag{6}$$

Alternative Approaches

$$Q(\pi) = \frac{1}{2}\left(\mathbb{E}\left[Y_i\left(\pi\left(X_i\right)\right)\right] - \mathbb{E}\left[Y_i\left(-\pi\left(X_i\right)\right)\right]\right) \tag{7}$$

$$\widehat{Q}(\pi) = \frac{1}{n}\sum_{i=1}^{n} \pi\left(X_i\right)\widehat{\Gamma}_i \tag{8}$$

What is $\widehat{\Gamma}_i$?
Zhao (2015) assume a randomized controlled trial and use $\widehat{\Gamma}_i = W_i\, Y_i \,/\, \mathbb{P}\left[W_i = 1\right]$, while Kitagawa and Tetenov (forthcoming) uses inverse-propensity weighting $\widehat{\Gamma}_i = W_i\, Y_i \,/\, \hat{e}_{W_i}(X_i)$. In an attempt to stabilize the weights, Beygelzimer et al (2009) introduce an "offset"

$$\widehat{\Gamma}_i = \frac{W_i}{\hat{e}_{W_i}(X_i)}\left(Y_i - \frac{\max\left\{Y_i\right\} + \min\left\{Y_i\right\}}{2}\right),$$

while Zhao et al (2015) go further and advocate

$$\widehat{\Gamma}_i = \frac{W_i}{\hat{e}_{W_i}(X_i)}\left(Y_i - \frac{\hat{\mu}_{+1}(X_i) + \hat{\mu}_{-1}(X_i)}{2}\right).$$

Alternative Approaches

Class of estimators:

$$\widehat{Q}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \pi\left(X_i\right) \widehat{\Gamma}_i \tag{9}$$

What is $\widehat{\Gamma}_i$? Athey/Wager:

$$\widehat{\Gamma}_i := \hat{\mu}_{+1}^{(-k(i))}(X_i) - \hat{\mu}_{-1}^{(-k(i))}(X_i) + W_i \, \frac{Y_i - \hat{\mu}_{W_i}^{(-k(i))}(X_i)}{\hat{e}_{W_i}^{(-k(i))}\left(X_i\right)}, \tag{10}$$

where $k(i) \in \{1, \ldots, K\}$ denotes the fold containing the i-th obs.

Optimizing Q and Estimating Impact

Class of estimators:

$$\widehat{Q}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \pi\left(X_i\right) \widehat{\Gamma}_i \tag{11}$$

Optimize $\widehat{Q}(\pi)$: use a classifier with labels $\text{sign}(\widehat{\Gamma}_i)$ and weights $|\widehat{\Gamma}_i|$. This step is treated as a constrained optimization problem, not estimation.

Issue: extending to multiple classes.

Evaluating benefit of policy: Re-estimate policy using leave one out only at second stage; or leave one out for whole routine. Assign using re-estimated policy. Do this at all 10 folds; use this overall assignment as a unified policy. Then, use ATE methods to estimate difference between estimated treatment effect for group assigned to treatment, and the estimated treatment effect for group assigned to control. That is overall estimated benefit of the policy.

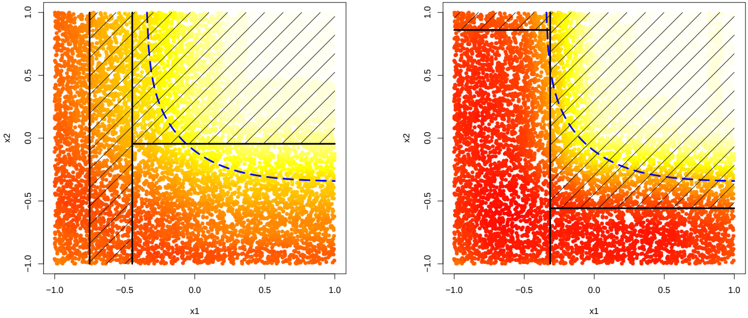
Athey/Wager Main Result

- Let $V(\pi)$ denote the semiparametrically efficient variance for estimating $Q(\pi)$.
- Let $V_* := V(\pi^*)$ denote the semiparametrically efficient variance for evaluating π^*
- Let $V_{\max}$ denote a sharp bound on the worst case efficient variance $\sup_\pi V(\pi)$ for any policy π. Results:
- Given policy class Π with VC dimension $\mathrm{VC}(\Pi)$, proposed learning rule yields policy $\hat{\pi}$ with regret bounded by

$$R(\hat{\pi}) = \mathcal{O}_P\left(\sqrt{V_* \log\left(\frac{V_{\max}}{V_*}\right)\frac{\mathrm{VC}(\Pi)}{n}}\right). \qquad (12)$$

- We also develop regret bounds for non-parametric policy classes Π with a bounded entropy integral, such as finite-depth decision trees.

Illustration



Inverse-propensity weighting   Double machine learning

Figure: Results from two attempts at learning a policy π by counterfactual risk minimization over depth-2 decision tree, with $\widehat{Q}$-estimators obtained by inverse-propensity weighting and double machine learning. Dashed blue line denotes the optimal decision rule (treat in the upper-right corner, do not treat elsewhere); solid black lines denote learned policies π (treat in the shaded regions, not elsewhere). Heat map depicts average over 200 simulations.

Conclusions

Contributions from causal inference and econometrics literature:

- Identification and estimation of causal effects
- Classical theory to yield asymptotically normal and centered confidence intervals
- Semiparametric efficiency theory

Contributions from ML:

- Practical, high performance algorithms for personalized prediction and policy estimation

Putting them together:

- Practical, high performance algorithms
- Causal effects with valid confidence intervals
- Consistent with insights from efficiency theory