Computational Data Analytics
for Economists

Lecture 4

Optimal Policy Learning

Helge Liebert   Anthony Strittmatter

Outline

1. Learning Policies from CATEs?
2. Optimal Policy Learning
3. Applications
4. Extensions

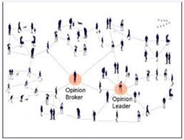
Literature

- Athey and Wager (2018): ”Efficient Policy Learning”, download.
- Kitagawa and Tetenov (2018): ”Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice”, Econometrica, 86(2), pp. 591-616, download.

Targeting Treatments

Possible Questions:

- Who should receive the offer to participate in a training program?
- How should we design the eligibility criteria of a welfare program?
- Who should receive a direct fund-raising mail to increase charitable giving?
- Who should be solicited during electoral campaigning?



⇒ Optimal policy rules can improve the allocation of limited resources, e.g., to save budget or to improve welfare

What are Policy Rules?

- Determine the allocation of treatments to individuals with different observable covariates
- Purpose is to find a policy rule π based on the covariates $X \rightarrow \{-1, 1\}$
- The policy rule $\pi(X_i)$ is 1 when individual i is assigned to the treatment and -1 otherwise
- Policy rules are often called assignment rule, individualised treatment rule (ITR), personalised treatment rule, etc.
- Policy rules are closely related to treatment effects (CATEs)
- But instead of estimating the effect, we want to learn an optimal policy rule

Notation

- $W_i = 1$ when individual i is treated and $W_i = -1$ otherwise
- Potential outcomes are $Y_i(w)$ for $w \in \{-1, 1\}$
- Observed outcome:

$$Y_i = Y_i(-1) + \frac{1 + W_i}{2} \left(Y_i(1) - Y_i(-1)\right)$$

- Outcome under policy rule $\pi(X_i)$:

$$Y_i(\pi(X_i))$$

- Individual causal effect:

$$\delta_i = Y_i(1) - Y_i(-1)$$

- CATEs:

$$\delta(x) = E[\delta_i | X_i = x]$$

Approach Based on CATEs

- Optimal policy rule:

$$\pi^* = \pi(\delta_i) = 1\{\delta_i > 0\} - 1\{\delta_i \leq 0\}$$

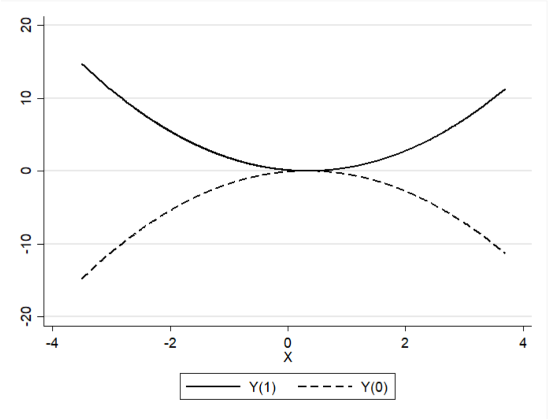
- CATE based policy rule:

$$\pi(\delta(X_i)) = 1\{\delta(X_i) > 0\} - 1\{\delta(X_i) \leq 0\}$$

- The selection of a policy rule is a classification problem
- CATEs are not targeted at this classification problem

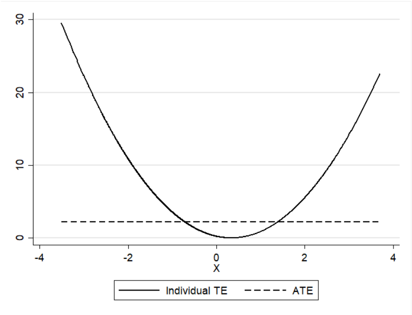
Simple Example

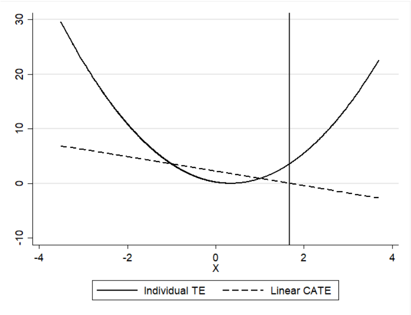
- $X \sim N(0,1)$
- $Y(1) = (X - 1/3)^2$
- $Y(0) = -(X - 1/3)^2$



Reference: Qian and Murphy (2011)

CATEs are Not Suited to Find Optimal Policies





- Treating everybody is optimal
- ATEs find optimal policy rule ($MSE_{ATE} = 9.4$), even though linear prediction of CATEs approximate the individual treatment effects better ($MSE_{ATE} > MSE_{CATE} = 7.8$)

Policy Learning Approach

- In the optimal case, we select $Y_i(1)$ when $\delta_i > 0$ and $Y_i(-1)$ when $\delta_i < 0$, such that
$$\pi^* = \max_{\pi} E[Y_i(\pi(X_i))]$$
- This is equivalent to selecting the π that minimises the regret function
$$R(\pi) = E[Y_i(\pi^*(X_i))] - E[Y_i(\pi(X_i))]$$
$\rightarrow$ minimax regret criterion (Manski, 2004)
- The regret is the gap between the optimal and estimated policy
- Type I regret: due to mistakenly choosing an inferior treatment
- Type II regret: due to mistakenly rejecting a superior treatment innovation

How Can We Do This in Practice?

- Define a policy value function

$$Q(\pi) = E[Y_i(\pi(X_i))]$$

or

$$\begin{aligned} Q(\pi) =& E[Y_i(\pi(X_i))] - \frac{1}{2}E[Y_i(1) + Y_i(-1)] \\ =& \frac{1}{2}E[\pi(X_i)\{Y_i(1) - Y_i(-1)\}] \end{aligned}$$

- Estimate the policy $\hat{\pi}(X_i)$ that maximises $Q(\pi)$

$$\hat{\pi} = \max_{\pi} \hat{Q}(\pi)$$

- **Main Challenge:** How to estimate $\hat{Q}(\pi)$? $\Rightarrow$ Q-learning

Modified Outcome Method

- Inverse Probability Weighting (IPW):

$$\hat{Q}(\pi) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1\{W_i = \pi(X_i)\}Y_i}{p_\pi(X_i)}\right]$$

with $p_\pi(X_i) = Pr(W_i = \pi(X_i)|X_i)$

- Augmented IPW:

$$\hat{Q}(\pi) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1\{W_i = \pi(X_i)\}(Y_i - \mu_\pi(X_i)}{p_\pi(X_i)} + \mu_\pi(X_i)\right]$$

with $\mu_\pi(X_i) = E[Y_i|W_i = \pi(X_i), X_i]$

- Maximization of these non-smooth functions is difficult

Reference: Zhang, Tsiatis, Laber, Davidian (2012)

Inverse Probability Weighting (IPW)

$$
\begin{aligned}
E[Q(\pi)|X_i = x] =& E\left[\left. \frac{1\{W_i = \pi(X_i)\}Y_i}{p_\pi(X_i)} \right| X_i\right] \\
\stackrel{LIE}{=}& E\left[\left. \frac{1\{W_i = \pi(X_i)\}Y_i}{p_\pi(X_i)} \right| W_i = \pi(X_i), X_i\right] p_\pi(X_i) \\
=& E\left[Y_i | W_i = \pi(X_i), X_i\right] \\
=& E\left[Y_i(\pi(X_i)) | W_i = \pi(X_i), X_i\right] \\
\stackrel{CIA}{=}& E\left[Y_i(\pi(X_i)) | X_i\right]
\end{aligned}
$$

Alternative Policy Value Function

- Define the alternative policy value function

$$
\begin{aligned}
Q(\pi) &= E[Y_i(\pi(X_i))] - E[Y_i(-\pi(X_i))] \\
&= E[\pi(X_i)\{Y_i(1) - Y_i(-1)\}]
\end{aligned}
$$

- Then

$$
\hat{Q}(\pi) = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}(X_i)\hat{\Gamma}_i
$$

- We can estimate the score $\hat{\Gamma}_i$ using sample splitting

Score Functions

- IPW:

$$\hat{\Gamma}_i = \frac{W_i}{\hat{p}_\pi(X_i)} Y_i$$

- Beygelzimer and Langford (2009) (”offset”):

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

- Zhao, Zeng, Laber, Song, Yuan, and Kosorok (2015):

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

- Athey and Wager (2018) (DML):

$$\hat{\Gamma}_i = \hat{\mu}_{+1}(X_i) - \hat{\mu}_{-1}(X_i) + W_i \frac{Y_i - \hat{\mu}_\pi(X_i)}{\hat{p}_\pi(X_i)}$$

Weighted Classification Representation

- Note that the policy value function

$$\hat{Q}(\pi) = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}(X_i)\hat{\Gamma}_i$$

  equals

$$\hat{Q}(\pi) = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}(X_i) sign(\hat{\Gamma}_i)|\hat{\Gamma}_i|$$

- Classification of $sign(\Gamma_i)$ with weights $|\Gamma_i|$
- **Intuitively:**
  - Misclassification hurts more when the (absolute) treatment effects are large
  - Misclassification of individuals with almost zero effects is not very costly

Reference: Zhao, Zeng, Rush, and Kosorok (2012)

Objective Function

- Zhang, Tsiatis, Davidian, Zhang, and Laber (2012) note that we can solve the weighted classicfication problem by

$$\min_{\pi} \sum_{i=1}^{N} |\Gamma_i| \left(1\{\Gamma_i > 0\} - \hat{\pi}(X_i)\right)^2$$

- We have to find some classification function for $\hat{\pi}(X_i)$ that minimises the objective function
- The estimated policy rule $\hat{\pi}(X_i)$ maximises $Q(\pi)$

Policy Learning Algorithm

1. Split the data in two samples A and B
2. Use ML to estimate $\hat{\mu}^A_{+1}(X_i)$, $\hat{\mu}^A_{+1}(X_i)$, and $\hat{p}^A_{+1}(X_i)$ in Sample A; as well as $\hat{\mu}^B_{+1}(X_i)$, $\hat{\mu}^B_{+1}(X_i)$, and $\hat{p}^B_{+1}(X_i)$ in Sample A
3. Estimate your preferred score function $\hat{\Gamma}_i$, for example,

$$\hat{\Gamma}^A_i = \hat{\mu}^B_{+1}(X_i) - \hat{\mu}^B_{-1}(X_i) + W_i \frac{Y_i - \hat{\mu}^B_{\pi}(X_i)}{\hat{p}^B_{\pi}(X_i)}$$

$$\hat{\Gamma}^B_i = \hat{\mu}^A_{+1}(X_i) - \hat{\mu}^A_{-1}(X_i) + W_i \frac{Y_i - \hat{\mu}^A_{\pi}(X_i)}{\hat{p}^A_{\pi}(X_i)}$$

4. Use ML to classify $sign(\hat{\Gamma}^A_i)$ with weight $|\hat{\Gamma}^A_i|$ in order to obtain the probability $\hat{q}^A_i(X_i) = Pr(\hat{\pi}^A(X_i) = 1)$. Proceed equivalently in sample B and obtain $\hat{q}^B_i(X_i)$.
5. Implement the policy rule $\pi(\hat{X}_i) = 2 \cdot 1\{\hat{q}^A_i(X_i) + \hat{q}^B_i(X_i) > 1\} - 1$

Classification Methods

- **Classification Trees**
  - In contrast to regression trees, classification trees use different performance measures
  - These measures are targeted to minimise the impurity (instead of the regression fit)
  - Entropy or Gini index
- **Logistic LASSO**
- **Support Vector Machines** (partition data in two samples)

Regret Bounds

- Kitagawa and Tetenov (2018):

$$R(\hat{\pi}) = \mathscr{O}_P\left(\frac{M}{\eta}\sqrt{\frac{VC(\Pi)}{N}}\right)$$

- Cortes, Mansour, and Mohri (2010) and Swaminathan and Joachims (2015):

$$R(\hat{\pi}) = \mathscr{O}_P\left(\sqrt{V^*\frac{log(N)VC(\Pi)}{N}}\right)$$

- Athey and Wager (2018):

$$R(\hat{\pi}) = \mathscr{O}_P\left(\sqrt{V^* log\left(\frac{V_{max}}{V^*}\right)\frac{VC(\Pi)}{N}}\right)$$

- $Y_i \leq M$ and $\eta \leq p_\pi(X_i) < 1-\eta$
- $V^* = V(\pi^*)$ is semiparametrically efficient variance for $Q(\pi)$
- V_{max} is sharp bound on worst case efficient variance $sup_\pi V(\pi)$
- $VC(\Pi)$ is policy class with Vapnik-Chervonenkis dimension

Regularity Conditions

Athey and Wager (2018):

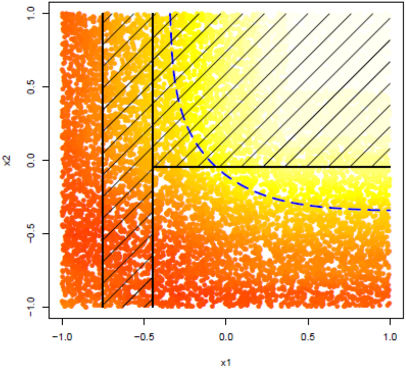
- Uniform consistency of $\hat{\mu}_\pi(X_i)$ and $\hat{p}_\pi(X_i)$
- $\sqrt[4]{N}$-convergence of $\hat{\mu}_\pi(X_i)$ and $\hat{p}_\pi(X_i)$
- $VC(\Pi)$ needs to have bounded complexity

Further Results:

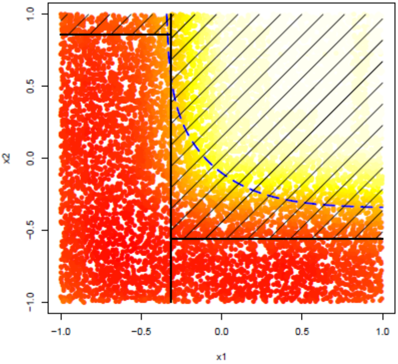
- $\hat{Q}_{DML}(\pi)$ is semi-parametrically efficient
- $\sqrt{N}(\hat{Q}_{DML}(\pi) - Q(\pi)) \xrightarrow{d} N(0, V(\pi))$
- $V(\pi)) = Var(\pi(X_i)\delta(X_i)) + E\left[\frac{Var(Y(-1)|X_i))}{\hat{p}_{-1}(X_i)} + \frac{Var(Y(+1)|X_i))}{\hat{p}_{+1}(X_i)}\right]$

Simulation Exercise

Inverse Probability Weighting



Double Machine Learning

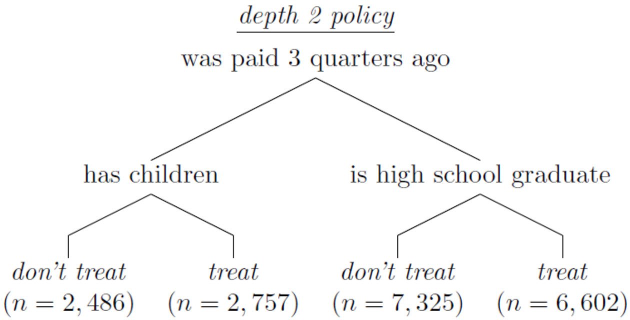


Application: California's GAIN program

- Welfare-to-work program that provides participants with a mix of educational resources and job search assistance MDRC conducted a randomised experiment
- Randomly chosen registrants were eligible to receive GAIN benefits immediately, whereas others were embargoed from the program
- Significant impact on earnings 9-years following the randomisation

$\Rightarrow$ **Question:** Can we to prioritize treatment to some subgroups of GAIN registrants who are particularly likely to benefit from it?

Application: California's GAIN program (cont.)



Source: Athey and Wager (2018)

Application: California's GAIN program (cont.)

Estimated Improvement (in 1,000 dollar)

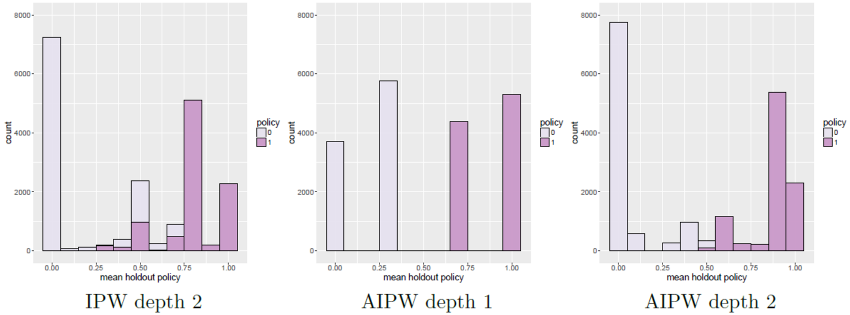
Method	Estimated Improvement	Width of Bound
causal forest	0.095	0.026
IPW depth 2	0.073	0.026
AIPW depth 1	0.065	0.026
AIPW depth 2	0.098	0.026

ATE = 0.141 to 0.208

Source: Athey and Wager (2018)

Comparison of In- and Out-of-Sample Policy Rules

We expect polarisation when in-sample and hold-out-sample rules align



Source: Athey and Wager (2018)

Budget constraints

- Substract cost (e.g., Kitagawa and Tetenov, 2018):

$$\hat{\Gamma}_i^{Budget} = \hat{\Gamma}_i - c_i$$

- Fix number of participants (e.g., Bhattacharya and Dupas, 2012):

$$\hat{\pi}_i^{Budget} = 2 \cdot 1\{\hat{\pi}(X_i) \geq \bar{\pi}\} - 1$$

- Combination of both enables to fix the number of participants when the cost of participation vary

Application: Job Corps

Outcome Variable:	*30-Month Post-Program Earnings, No Treatment Cost*		*30-Month Post-Program Earnings, $774 Cost for Each Assigned Treatment*	
Treatment Rule:	Share of Population to Be Treated	Est. Welfare Gain per Population Member	Share of Population to Be Treated	Est. Welfare Gain per Population Member
Treat everyone	1	$1,180 ($464, $1,896)	1	$404 (−$313, $1,121)
EWM quadrant rule	0.95	$1,340 ($441, $2,239)	0.8	$643 (−$258, $1,544)
EWM linear rule	0.96	$1,364 ($398, $2,330)	0.69	$792 (−$177, $1,761)
EWM linear rule (with (education)^2 and (education)^3)	0.88	$1,489 ($374, $2,603)	0.75	$897 (−$214, $2,008)
Linear regression plug-in rule	0.98	$1,152	0.86	$527
Linear regression plug-in rule (with (education)^2 and (education)^3)	0.95	$1,263	0.91	$547
Nonparametric plug-in rule	0.91	$1,693	0.78	$996

Source: Kitagawa and Tetenov (2018)

Batch vs. Bandit Algorithms

Batch

- Historical dataset
- Potentially optimal policy rules change over time
- Then findings cannot be extrapolated to the future

Bandit

- Data arrives sequentially (typically online data)
- Treatment decisions are made sequentially
- Presence of exploration vs. exploitation trade-off
- Example: targeted online advertisements
- Reference: Dimakopoulou, Zhou, Athey, and Imbens (2018)

Further Extensions

- **Multiple treatments**
  (e.g., Frölich, 2008, Kallus, 2017, Zhou, Athey, Wager, 2018)
- **Ordered treatments**
  (e.g., Chen, Fu, He, Kosorok, and Liu, 2018)
- **Dynamic treatments**
  (e.g., Zhang and Zhang, 2018, Zhao, Zheng, Laber, and Kosorok, 2015)
- **Continuous treatments**
  (e.g., Chen, Zheng, and Kosorok, 2016, Athey and Wager, 2018)

Ethical Concerns?

- Statistical discrimination even if we omit critical variable (e.g., gender, migration, etc.)
- Examples: hiring decisions, flight prices, program assignments
- More or less than discrimination than humans?
- Targeting rules also have the potential to reduce discrimination, but it has to be used appropriately
- Current scandals: Cambridge Analytica