Machine Learning Estimation of Heterogeneous Treatment Effects with Instruments

V. Syrgkanis¹, V. Lei², M. Oprescu¹, M. Hei¹, K. Battocchi¹, G. Lewis¹

¹Microsoft Research, New England (ALICE Project)
²TripAdvisor
TripAdvisor Membership Problem

- What is the causal effect of becoming a member on TripAdvisor on downstream activity on the webpage?
- How does that effect vary with observable characteristics of the user?
- Useful for understanding the quality of membership offering/improvements/targeting
TripAdvisor Membership Problem

- What is the causal effect of becoming a member on TripAdvisor on downstream activity on the webpage?
- How does that effect vary with observable characteristics of the user?
- Useful for understanding the quality of membership offering/improvements/targeting

Standard approach: Let’s run an A/B test!
Not applicable: We cannot enforce the treatment!
- We cannot take a random half of the users and make them members
- Membership is an action that requires user engagement!
Recommendation A/B Tests

- In optimizing a service we want to understand the causal effects of actions that involve user engagement (e.g. becoming a member)
Recommendation A/B Tests

- In optimizing a service we want to understand the causal effects of actions that involve user engagement (e.g. becoming a member)

- We can run a recommendation A/B test:
  - “recommend/create extra incentives” to half the users to take the action/treatment
  - *Example at TripAdvisor:* enable an easier sign-up flow process for a random half of users
In optimizing a service we want to understand the causal effects of actions that involve user engagement (e.g. becoming a member).

- We can run a recommendation A/B test:
  - “recommend/create extra incentives” to half the users to take the action/treatment
  - *Example at TripAdvisor:* enable an easier sign-up flow process for a random half of users

- **Non-Compliance:** “user’s choice to comply or not” can lead to biased estimates
Instrumental Variables (IV)

- **Instrumental Variable**: any random variable $Z$ that affects the treatment assignment $T$ but does not affect the outcome $Y$ other than through the treatment.
- Cohort assignment in recommendation A/B test is an instrument.
- We can apply IV methods to estimate average treatment effect $\theta$. 
Instrumental Variables (IV)

- **Instrumental Variable**: any random variable $Z$ that affects the treatment assignment $T$ but does not affect the outcome $Y$ other than through the treatment.
- Cohort assignment in recommendation A/B test is an instrument.
- We can apply IV methods to estimate average treatment effect $\theta$.

- Typical IV methods do not account for complex effect or compliance heterogeneity.
This Work: Personalized/Heterogeneous Effects

Personalization requires estimates of heterogeneous effect $\theta(X)$ as a function of observable characteristics $X$. 
This Work: Personalized/Heterogeneous Effects

Personalization requires estimates of *heterogeneous effect* $\theta(X)$ as a function of observable characteristics $X$

◊ Can we learn complex/non-linear models for the heterogeneous effect $\theta(X)$?

◊ Can we reduce estimation to standard ML problems like regression/classification?
Reducing to Regression/Classification

- Consider the **compliance score** (Abadie’03)

\[
\Delta(X) = (2Z - 1) \frac{P(T = 1|Z = 1, X) - P(T = 1|Z = 0, X)}{2}
\]

- Let \( \tilde{Y} = Y - \mathbb{E}[Y|X] \) and \( \tilde{T} = T - \mathbb{E}[T|X] \)

- Estimate preliminary \( \hat{\theta}(X) \)

\[
\hat{\theta} = \arg\min_{\theta} \mathbb{E} \left[ \left( \tilde{Y} - \theta(X) \cdot \Delta(X) \right)^2 \right]
\]

- Estimate **robust final** \( \theta(X) \)

\[
\min_{\theta} \mathbb{E} \left[ \left( \hat{\theta}(X) + \frac{\tilde{Y} - \hat{\theta}(X) \cdot \tilde{T}}{\Delta(X)} - \theta(X) \right)^2 \right]
\]
Reducing to Regression/Classification

- Consider the **compliance score** (Abadie'03)
  \[ \Delta(X) = \frac{\Pr(T = 1|Z = 1, X) - \Pr(T = 1|Z = 0, X)}{2} \]

- Let \( \tilde{Y} = Y - \mathbb{E}[Y|X] \) and \( \tilde{T} = T - \mathbb{E}[T|X] \)

- Estimate **preliminary** \( \hat{\theta}(X) \)

- Estimate **robust final** \( \theta(X) \)

\[ \hat{\theta} = \arg\min_{\theta(\cdot)} \mathbb{E} \left[ \left( \tilde{Y} - \theta(X) \cdot \Delta(X) \right)^2 \right] \]

\[ \min_{\theta(\cdot)} \mathbb{E} \left[ \left( \hat{\theta}(X) + \frac{\tilde{Y} - \hat{\theta}(X) \cdot \tilde{T}}{\Delta(X)} - \theta(X) \right)^2 \right] \]
Reducing to Regression/Classification

Classification

- Consider the compliance score (Abadie ’03)
  \[
  \Delta(X) = \frac{2(Z - 1) - \frac{\mathbb{P}(T = 1|Z = 1, X) - \mathbb{P}(T = 1|Z = 0, X)}{2}}
  \]
- Let \( \bar{Y} = Y - \mathbb{E}[Y|X] \) and \( \bar{T} = T - \mathbb{E}[T|X] \)
- Estimate preliminary \( \hat{\theta}(X) \)

Regression

- \( \hat{\theta} = \arg\min_{\theta(X)} \mathbb{E}\left[\left(\bar{Y} - \theta(X) \cdot \Delta(X)\right)^2\right] \)
- Estimate robust final \( \theta(X) \)

\[
\min_{\theta(X)} \mathbb{E}\left[\left(\hat{\theta}(X) + \frac{\bar{Y} - \hat{\theta}(X) \cdot \bar{T}}{\Delta(X)} - \theta(X)\right)^2\right]
\]

Benefits of Reduction Approach

- **Statistical and computational benefits** of modern ML approaches (forests, regularized linear models, SVM, DNNs etc.)

- **Cross-validation** for model selection and hyperparameter tuning

- **Interpretability** of estimated models (SHAP, Lime, Influence functions)
MSE
Robustness

- Loss function for final estimate satisfies Neyman orthogonality [Chernozhukov et al.’16, Foster – Syrgkanis’19]

- Mean-Squared-Error of final $\theta(X)$ robust to errors in auxiliary Classifications/Regressions
MSE
Robustness

- Loss function for final estimate satisfies Neyman orthogonality [Chernozhukov et al.’16, Foster – Syrgkanis’19]

- Mean-Squared-Error of final $\theta(X)$ robust to errors in auxiliary Classifications/Regressions

- Approach extends beyond recommendation A/B tests, to linear-in-treatment IV setting

- Resolves open question in literature [Nie-Wager’17]
Confidence Intervals (CIs)

- When final regression supports CI construction, Neyman orthogonality typically preserves the validity of the intervals
  - Inference on best linear projection of heterogeneous effect via OLS
  - Inference on high-dimensional linear projections via Debiased Lasso
  - Non-Parametric inference via Honest Regression Forests
TripAdvisor Experiment

For random half of 4 million users, easier sign-up flow was enabled
- Easier sign-up incentivizes membership
- Outcome: number of visits in the next 14 days

High Level Take-Aways
- Large heterogeneity based on which pages were recently visited
- Large heterogeneity based on platform of access (e.g. iPhone, Linux etc.)
- Results enable better targeting of right user population and improvements of membership offering for user segments with small/almost zero effects
Try it Out and Check out Poster #185!

- Code: [https://github.com/microsoft/EconML/tree/master/prototypes/dml_iv](https://github.com/microsoft/EconML/tree/master/prototypes/dml_iv)

```python
dr_cate = IntentToTreatDRIV(model_y_x=RandomForestRegressor(),
                           model_t_xz=RandomForestClassifier(),
                           prel_model_effect=RandomForestRegressor(),
                           final_model_effect=LinearRegression())

dr_cate.fit(y, T, X, Z)
dr_cate.effect(X)
```

**EconML** python library for ML Estimation of Heterogeneous Treatment Effects

- [https://github.com/microsoft/EconML](https://github.com/microsoft/EconML)
- `pip install econml`

**ALICE** (Automated Learning and Intelligence for Causation and Economics) project: