High-Dimensional Causal Inference

Kosuke Imai

Department of Politics
Center for Statistics and Machine Learning
Princeton University

Talk at Machine Learning for Causal Inference Workshop

Harvard and MIT
February 12, 2016

Joint work with Naoki Egami

High-Dimensional Causal Inference

1. Treatment effect heterogeneity
   - How do treatment effects vary across individuals?
   - Who benefits from (or is harmed by) the treatment?
2. Treatment heterogeneity
   - What aspects of a treatment are responsible for causal effects?
   - What combination of treatments is efficacious?
3. Individualized treatment regimes
   - What combination of treatments is optimal for a given individual?

- Causal prediction vs. causal learning in high-dimension
- Importance of interpretability in high-dimension

Factorial Designs with Many Treatments

- The most basic form of high-dimensional causal inference
- Many treatments, each having multiple levels
- A motivating application: Conjoint analysis (Hainmueller *et al.* 2014)
  - survey experiments to measure immigration preferences
  - a representative sample of 1,396 American adults
  - each respondent evaluates 5 pairs of immigrant profiles
  - $\texttt{gender}^2$, $\texttt{education}^7$, $\texttt{origin}^{10}$, $\texttt{experience}^4$, $\texttt{plan}^4$, $\texttt{language}^4$, $\texttt{profession}^{11}$, $\texttt{application reason}^3$, $\texttt{prior trips}^5$
  - Over 1 **million** treatment combinations!
  - What combinations of immigrant characteristics make them preferred?
- Too many treatment combinations $\rightsquigarrow$ Need for an effective summary
- Many potential applications in academia and industry

Machine Learning and Causal Inference

- How should we analyze the data from a factorial randomized experiment with many treatments?
- Regression model: $\mathbb{E}(Y \mid \mathbf{T}) = f(\mathbf{T})$
- There are many machine learning methods to estimate this model
- In this setting, causal inference is a prediction problem $\mathbb{E}(Y(\mathbf{t})) = \mathbb{E}(Y \mid \mathbf{T} = \mathbf{t}) = f(\mathbf{t})$
- But, how do we interpret these models?
- Scientists wish to understand the causal structure
  1. Predict $\mathbb{E}(Y(\mathbf{t}))$ using each treatment combination $\mathbf{t}$ and look at what values of $\mathbf{t}$ yield high/low predicted values of Y
     $\rightsquigarrow$ Finding patterns is difficult in high dimension
  2. Use a sparse regression model
     $\rightsquigarrow$ Difficult to interpret interaction terms (lack of invariance to the baseline condition)
- Causal testing vs. causal exploration

Causal Effects with Two Multi-valued Treatment Variables

1. Average Treatment Combination Effect (ATCE):
   - Average effect of treatment combination $(A, B) = (a_\ell, b_m)$ relative to the baseline condition $(A, B) = (a_0, b_0)$

   $$\tau(a_\ell, b_m; a_0, b_0) \equiv \mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_0)\}$$

   - Which treatment combination is most efficacious?

2. Average Marginal Treatment Effect (AMTE; Hainmueller et al. 2014):
   - Average effect of treatment $A = a_\ell$ relative to the baseline condition $A = a_0$ averaging over the other treatment B

   $$\psi(a_\ell, a_0) \equiv \int_{\mathcal{B}} \mathbb{E}\{Y(a_\ell, B) - Y(a_0, B)\} dF(B)$$

   - Which treatment is effective on average?

Other treatments can be integrated out

3. Average Marginal Treatment Interaction Effect (**AMTIE**):

$$\pi(a_\ell, b_m; a_0, b_0) \equiv \underbrace{\tau(a_\ell, b_m; a_0, b_0)}_{\text{ATCE of } (A,B)=(a_\ell, b_m)} - \underbrace{\psi(a_\ell, a_0)}_{\text{AMTE of } a_\ell} - \underbrace{\psi(b_m, b_0)}_{\text{AMTE of } b_m}$$

- Additional effect induced by $A = a_\ell$ and $B = b_m$ together beyond the separate effect of $A = a_\ell$ and that of $B = b_m$
- Unlike the standard interaction effects, the **AMTIE**s are invariant to the choice of baseline category
- However, the **AMTE**s and **AMTIE**s do depend on the distribution of treatment assignment
- Two solutions:
  1. use the treatment assignment probabilities from the experiment
  2. use the distribution of treatments in the target population

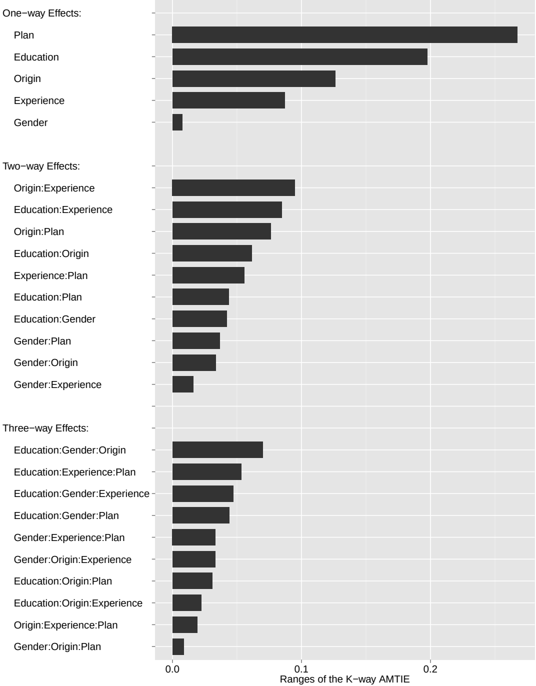
Causal Interaction in High-Dimension

- Definition: the difference between the ATCE and the sum of lower-order **AMTIE**s

- Example: 3-way **AMTIE**, $\pi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})$, equals

$$\underbrace{\tau_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})}_{\text{ATCE}}$$
$$- \underbrace{\left\{\pi_{1:2}(t_1, t_2; t_{01}, t_{02}) + \pi_{2:3}(t_2, t_3; t_{02}, t_{03}) + \pi_{1:3}(t_1, t_3; t_{01}, t_{03})\right\}}_{\text{sum of 2-way AMTIEs}}$$
$$- \underbrace{\left\{\psi(t_1; t_{01}) + \psi(t_2; t_{02}) + \psi(t_3; t_{03})\right\}}_{\text{sum of (1-way) AMTEs}}$$

- Contrast this with the standard higher-order interaction:
  3-way interaction effect = difference between 2-way interaction effects



- Sparse regression with one-way, two-way, and three-way effects
- Range of **AMTIE**s: importance of each factor and factor interaction
- Sparcity-of-effects principle
- `gender` appears to play a significant role in three-way interactions

3-way Interaction (education $\times$ gender $\times$ origin)

$$\underbrace{\tau(\texttt{Graduate}, \texttt{Male}, \texttt{India}; \texttt{Graduate}, \texttt{Female}, \texttt{India})}_{7.46} \qquad (n = 52; \quad n = 40)$$

$$= \underbrace{\psi(\texttt{Male}; \texttt{Female})}_{-0.77} + \underbrace{\pi(\texttt{Graduate}, \texttt{Male}; \texttt{Graduate}, \texttt{Female})}_{-0.34}$$

$$+ \underbrace{\pi(\texttt{Male}, \texttt{India}; \texttt{Female}, \texttt{India})}_{1.56} + \underbrace{\pi(\texttt{Graduate}, \texttt{Male}, \texttt{India}; \texttt{Graduate}, \texttt{Female}, \texttt{India})}_{7.01}$$

$$\underbrace{\tau(\texttt{High school, Male, Germany; High school, Female, Germany})}_{-11.52}$$

$(n = 41; \quad n = 56)$

$$= \underbrace{\psi(\texttt{Male}; \texttt{Female})}_{-0.77} + \underbrace{\pi(\texttt{High school, Male; High school, Female})}_{-0.67}$$

$$+ \underbrace{\pi(\texttt{Male}, \texttt{Germany}; \texttt{Female}, \texttt{Germany})}_{-3.34}$$

$$+ \underbrace{\pi(\texttt{High school, Male,Germany; High school, Female,Germany})}_{-6.74}.$$

Challenges of High-dimensional Causal Inference

- Estimation and inference $\rightsquigarrow$ machine learning and statistics
- Interpretation $\rightsquigarrow$ causal inference
- Experimental design
  - Multi-armed bandits in high-dimension
  - More noise $\rightsquigarrow$ sensitivity to the choice of tuning parameter
  - Linear UCB with variable selection $\rightsquigarrow$ attains oracle properties
  - Issues of dynamic variable selection in high-dimension