Causal Interaction in High Dimension

Naoki Egami Kosuke Imai

Princeton University

Joint Statistical Meetings

August 12, 2015

Causal Interaction in High Dimension

- Causal interaction: what treatment combinations are effective?
- High dimension = 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
- Interaction effects play an essential role

Two Interpretations of Causal Interaction

- Two binary treatments: A and B
- Potential outcomes: $Y(a, b)$ where $a, b \in \{0, 1\}$
- Conditional effect interpretation:

$$\underbrace{[Y(1,1) - Y(0,1)]}_{\text{effect of } A \text{ when } B=1} - \underbrace{[Y(1,0) - Y(0,0)]}_{\text{effect of } A \text{ when } B=0}$$

$\rightsquigarrow$ requires the specification of moderator

- Interactive effect interpretation:

$$\underbrace{[Y(1,1) - Y(0,0)]}_{\text{effect of } A \text{ and } B} - \underbrace{[Y(1,0) - Y(0,0)]}_{\text{effect of } A \text{ when } B=0} - \underbrace{[Y(0,1) - Y(0,0)]}_{\text{effect of } B \text{ when } A=0}$$

$\rightsquigarrow$ requires the specification of baseline condition

- The same quantity but two different interpretations

Difficulty of the Conventional Approach

- Lack of invariance to the baseline condition
  $\rightsquigarrow$ Inference depends on the choice of baseline condition

- 3×3 example:
  - Treatment $A \in \{a_0, a_1, a_2\}$ and Treatment $B \in \{b_0, b_1, b_2\}$
  - Regression model with the baseline condition (a_0, b_0):

$$\mathbb{E}(Y \mid A, B) \quad = \quad 1 + a_1^* + a_2^* + b_2^* + a_1^* b_2^* + 2a_2^* b_2^* + 3a_2^* b_1^*$$

  - Interaction effect for (a_2, b_2) $>$ Interaction effect for (a_1, b_2)

  - Another equivalent model with the baseline condition (a_0, b_1):

$$\mathbb{E}(Y \mid A, B) \quad = \quad 1 + a_1^* + 4a_2^* + b_2^* + a_1^* b_2^* - a_2^* b_2^* - 3a_2^* b_0^*$$

  - Interaction effect for (a_2, b_2) $<$ Interaction effect for (a_1, b_2)
  - Interaction effect for (a_2, b_1) is zero under the second model
  - All interaction effects with at least one baseline value are zero

Empirical Illustration: Lack of Invariance

- Linear regression with main effects and two-way interactions
- Baseline: *lowest* levels of job experiences and education

Job experience	**Education**						
	None	4th grade	8th grade	High school	Two-year college	College	Graduate
None	0	0	0	0	0	0	0
1–2 years	0	0.009	−0.019	−0.032	0.100	−0.044	−0.064
3–5 years	0	0.016	0.056	0.165	0.107	0.010	0.117
>5 years	0	−0.050	0.126	0.042	0.058	−0.094	0.015

The Effects of Changing the Baseline Condition

- Same linear regression but different baseline
- Baseline: *highest* levels of job experiences and education

	Education						
Job experience	None	4th grade	8th grade	High school	Two-year college	College	Graduate
None	0.015	0.065	−0.111	−0.027	−0.043	0.109	0
1–2 years	0.078	0.138	−0.066	0.006	0.120	0.129	0
3–5 years	−0.102	−0.036	−0.172	0.021	−0.054	0.002	0
>5 years	0	0	0	0	0	0	0

The Contributions of the Paper

1. Problems of the conventional approach:
   - Lack of invariance to the choice of baseline condition
   - Difficulty of interpretation for higher-order interaction
2. Solution: Average Marginal Treatment Interaction Effect
   - invariant to baseline condition
   - same, intuitive interpretation even for high dimension
   - simple estimation procedure
3. Reanalysis of the immigration survey experiment

Two-way Causal Interaction

- Two factorial treatments:

$$A \in \mathcal{A} = \{a_0, a_1, \ldots, a_{D_A-1}\}$$
$$B \in \mathcal{B} = \{b_0, b_1, \ldots, b_{D_B-1}\}$$

- Assumption: Full factorial design
  1. Randomization of treatment assignment

$$\{Y(a_\ell, b_m)\}_{a_\ell \in \mathcal{A}, b_m \in \mathcal{B}} \perp\!\!\!\perp \{A, B\}$$

  2. Non-zero probability for all treatment combination

$$\Pr(A = a_\ell, B = b_m) > 0 \quad \text{for all } a_\ell \in \mathcal{A} \quad \text{and} \quad b_m \in \mathcal{B}$$

- Fractional factorial design not allowed
  1. Use a small non-zero assignment probability
  2. Focus on a subsample
  3. Combine treatments

Non-Interaction Effects of Interest

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?

The Conventional Approach to Causal Interaction

- Average Treatment Interaction Effect (ATIE):

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

- Conditional effect interpretation:

$$\underbrace{\mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_m)\}}_{\text{Effect of } A = a_\ell \text{ when } B = b_m} - \underbrace{\mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\}}_{\text{Effect of } A = a_\ell \text{ when } B = b_0}$$

- Interactive effect interpretation:

$$\underbrace{\tau(a_\ell, b_m; a_0, b_0)}_{\text{ATCE}} - \underbrace{\mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\}}_{\text{Effect of } A = a_\ell \text{ when } B = b_0} - \underbrace{\mathbb{E}\{Y(a_0, b_m) - Y(a_0, b_0)\}}_{\text{Effect of } B = b_m \text{ when } A = a_0}$$

- Estimation: Linear regression with interaction terms

Lack of Invariance to the Baseline Condition

- Comparison between two ATIEs should not be affected by the choice of baseline conditions
- We prove that the ATIEs are *neither interval or order invariant*

- Interval invariance:

$$\begin{aligned} & \xi(a_\ell, b_m; a_0, b_0) \; - \; \xi(a_{\ell'}, b_{m'}; a_0, b_0) \\ = \; & \xi(a_\ell, b_m; a_{\tilde{\ell}}, b_{\tilde{m}}) \; - \; \xi(a_{\ell'}, b_{m'}; a_{\tilde{\ell}}, b_{\tilde{m}}), \end{aligned}$$

- Order invariance:

$$\begin{aligned} & \xi(a_\ell, b_m; a_0, b_0) \; \geq \; \xi(a_{\ell'}, b_{m'}; a_0, b_0) \\ \Longleftrightarrow \; & \xi(a_\ell, b_m; a_{\tilde{\ell}}, b_{\tilde{m}}) \; \geq \; \xi(a_{\ell'}, b_{m'}; a_{\tilde{\ell}}, b_{\tilde{m}}). \end{aligned}$$

The New Causal Interaction Effect

- Average Marginal Treatment Interaction Effect (**AMTIE**):

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

- Interactive effect interpretation: additional effect induced by $A = a_\ell$ and $B = b_m$ together beyond the separate effect of $A = a_\ell$ and that of $B = b_m$
- Compare this with ATIE:

$$\underbrace{\tau(a_\ell, b_m; a_0, b_0)}_{\text{ATCE}} - \underbrace{\mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\}}_{\text{Effect of } A=a_\ell \text{ when } B=b_0} - \underbrace{\mathbb{E}\{Y(a_0, b_m) - Y(a_0, b_0)\}}_{\text{Effect of } B=b_m \text{ when } A=a_0}$$

- We prove that the **AMTIE**s are both *interval and order invariant*
- The **AMTIE**s do depend on the distribution of treatment assignment
  1. specified by one's experimental design
  2. motivated by the target population

AMTIE is Invariant to the Choice of Baseline Condition

Job experience	**Education** None	4th grade	8th grade	High school	Two-year college	College	Graduate
None	0	−0.004	−0.028	−0.035	−0.031	0.012	−0.010
1–2 years	−0.001	−0.001	−0.025	−0.040	0.024	−0.009	−0.044
3–5 years	−0.040	−0.019	−0.042	0.031	−0.026	−0.022	0.024
$>$ 5 years	−0.014	−0.031	0.041	−0.011	−0.021	−0.036	−0.024

AMTIE is Invariant to the Choice of Baseline Condition

Job experience	**Education** None	4th grade	8th grade	High school	Two-year college	College	Graduate
None	0.024	0.020	−0.004	−0.011	−0.007	0.036	0.014
1–2 years	0.023	0.023	−0.001	−0.016	0.048	0.015	−0.020
3–5 years	−0.016	0.005	−0.018	0.055	−0.002	0.002	0.048
$>$ 5 years	0.010	−0.007	0.065	0.013	0.003	−0.012	0

The Relationships between the ATIE and the AMTIE

1. The **AMTIE** is a linear function of the ATIEs:

$$\begin{aligned}\pi(a_\ell, b_m; a_0, b_0) &= \xi(a_\ell, b_m; a_0, b_0) - \sum_{a \in \mathcal{A}} \Pr(A_i = a)\, \xi(a, b_m; a_0, b_0) \\ &\quad - \sum_{b \in \mathcal{B}} \Pr(B_i = b)\, \xi(a_\ell, b; a_0, b_0)\end{aligned}$$

2. The ATIE is also a linear function of the **AMTIE**s:

$$\xi(a_\ell, b_m; a_0, b_0) = \pi(a_\ell, b_m; a_0, b_0) - \pi(a_\ell, b_0; a_0, b_0) - \pi(a_0, b_m; a_0, b_0)$$

- Absence of causal interaction:
  All of the **AMTIE**s are zero if and only if all of the ATIEs are zero
- The **AMTIE**s can be estimated by first estimating the ATIEs

Higher-order Causal Interaction

- J factorial treatments: $\mathbf{T} = (T_1, \ldots, T_J)$
- Assumptions:
  1. Full factorial design

$$Y(\mathbf{t}) \quad \perp\!\!\!\perp \quad \mathbf{T} \quad \text{and} \quad \Pr(\mathbf{T} = \mathbf{t}) \ > \ 0 \quad \text{for all } \mathbf{t}$$

  2. Independent treatment assignment

$$T_j \quad \perp\!\!\!\perp \quad \mathbf{T}_{-j} \quad \text{for all } j$$

- Assumption 2 is not necessary for identification but considerably simplifies estimation

- We are interested in the K-way interaction where $K \leq J$
- We extend all the results for the 2-way interaction to this general case

Difficulty of Interpreting the Higher-order ATIE

- Generalize the 2-way ATIE by marginalizing the other treatments $\underline{\mathbf{T}}^{1:2}$

$$\xi_{1:2}(t_1, t_2; t_{01}, t_{02}) \equiv \int \mathbb{E}\left\{ Y(t_1, t_2, \underline{\mathbf{T}}^{1:2}) - Y(t_{01}, t_2, \underline{\mathbf{T}}^{1:2}) - Y(t_1, t_{02}, \underline{\mathbf{T}}^{1:2}) + Y(t_{01}, t_{02}, \underline{\mathbf{T}}^{1:2}) \right\} dF(\underline{\mathbf{T}}^{1:2})$$

- In the literature, the 3-way ATIE is defined as

$$\xi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03}) \equiv \underbrace{\xi_{1:2}(t_1, t_2; t_{01}, t_{02} \mid T_3 = t_3)}_{\text{2-way ATIE when } T_3 = t_3} - \underbrace{\xi_{1:2}(t_1, t_2; t_{01}, t_{02} \mid T_3 = t_{03})}_{\text{2-way ATIE when } T_3 = t_{03}}$$

- Higher-order ATIEs are similarly defined sequentially
- This representation is based on the conditional effect interpretation
- Problem: the conditional effect of conditional effects!

Interactive Effect Interpretation of the Higher-order ATIE

- We show that the higher-order ATIE also has an interactive effect interpretation
- Example: 3-way ATIE, $\xi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})$, equals

$$\underbrace{\tau_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})}_{\text{ATCE}}$$
$$- \{\xi_{1:2}(t_1, t_2; t_{01}, t_{02} \mid T_3 = t_{03}) + \xi_{2:3}(t_2, t_3; t_{02}, t_{03} \mid T_1 = t_{01})$$
$$+ \xi_{1:3}(t_1, t_3; t_{01}, t_{03} \mid T_2 = t_{02})\} \quad \text{sum of 2-way conditional ATIEs}$$
$$- \{\tau_1(t_1, t_{02}, t_{03}; t_{01}, t_{02}, t_{03}) + \tau_2(t_{01}, t_2, t_{03}; t_{01}, t_{02}, t_{03})$$
$$+ \tau_3(t_{01}, t_{02}, t_3; t_{01}, t_{02}, t_{03})\} \quad \text{sum of (1-way) ATCEs}$$

- Problems:
  1. Lower-order *conditional* ATIEs rather than lower-order ATIEs are used
  2. K-way ATCE $\neq$ sum of all K-way and lower-order ATIEs
  3. (We prove) Lack of invariance to the baseline conditions

The K-way Average Marginal Treatment Interaction Effect

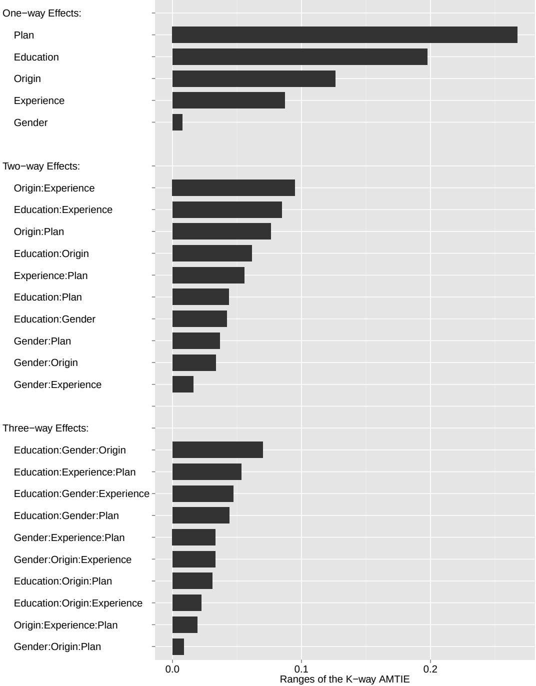
- Definition: the difference between the ATCE and the sum of lower-order **AMTIE**s
- Interactive effect interpretation
- Example: 3-way **AMTIE**, $\pi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})$, equals

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

- Properties:
  1. K-way ATCE $=$ the sum of all K-way and lower-order **AMTIE**s
  2. Interval and order invariance to the baseline condition
  3. Derive the relationships between the **AMTIE**s and ATIEs for any order

Empirical Analysis of the Immigration Survey Experiment

- 5 factors (`gender`^{2}, `education`^{7}, `origin`^{10}, `experience`^{4}, `plan`^{4})
  1. full factorial design assumption
  2. computational tractability
- Matched-pair conjoint analysis: randomly choose one profile
- Binary outcome: whether a profile is selected
- Model with one-way, two-way, and three-way interaction terms
- $p = 1,575$ and $n = 6,980$
- Curse of dimensionality $\Longrightarrow$ sparcity assumption
- Support vector machine with a lasso constraint (Imai & Ratkovic, 2013)
- Under-identified model that includes baseline conditions
- 99 non-zero and $1,476$ zero coefficients
- Cross-validation for selecting a tuning parameter
- `FindIt: Finding heterogeneous treatment 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

Decomposing the Average Treatment Combination Effect

- Two-way effect example (`origin` × `experience`):

$$\underbrace{\tau(\texttt{Somalia, 1-2 years; India, None})}_{-3.74} \qquad (n = 168; \quad n = 155)$$

$$= \underbrace{\psi(\texttt{Somalia}; \texttt{India})}_{-5.14} + \underbrace{\psi(\texttt{1} - \texttt{2years}; \texttt{None})}_{5.12} + \underbrace{\pi(\texttt{Somalia}, \texttt{1} - \texttt{2years}; \texttt{India}, \texttt{None})}_{-3.72}$$

- Three-way examples (`education` × `gender` × `origin`):

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

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

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

Concluding Remarks

- Interaction effects play an essential role in causal heterogeneity
  1. moderation
  2. causal interaction
- Two interpretations of causal interaction
  1. conditional effect interpretation (problematic in high dimension)
  2. interactive effect interpretation
- Average Marginal Treatment Interaction Effect
  1. interactive effect in high-dimension
  2. invariant to baseline condition
  3. enables effect decomposition
  4. $\rightsquigarrow$ effective analysis of interactions in high-dimension
- Estimation challenges in high dimension
  1. group lasso, hierarchical interaction
  2. post-selection inference