Causal Interaction in High Dimension

Naoki Egami          Kosuke Imai
Princeton University

Talk at Asian Political Methodology Conference
Tsingha University, Beijing
January 8, 2016
Interaction Effects and Causal Heterogeneity

1. Moderation
   - How do treatment effects vary across individuals?
   - Who benefits from (or is harmed by) the treatment?
   - Interaction between treatment and pre-treatment covariates

2. Causal interaction
   - What aspects of a treatment are responsible for causal effects?
   - What combination of treatments is efficacious?
   - Interaction between treatment variables

3. Individualized treatment regimes
   - What combination of treatments is optimal for a given individual?
Causal Interaction in High Dimension

- 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
  - gender\(^2\), education\(^7\), origin\(^{10}\), experience\(^4\), plan\(^4\), language\(^4\), profession\(^{11}\), application reason\(^3\), prior trips\(^5\)
  - Over 1 million treatment combinations!

  - What combinations of immigrant characteristics make them preferred?

  - Too many treatment combinations \(\implies\) Need for an effective summary
  - Interaction effects play an essential role
Two Interpretations of Causal Interaction

1 Conditional effect interpretation:
   - Does the effect of one treatment change as we vary the value of another treatment?
   - Does the effect of being black change depending on whether an applicant is male or female?
   - Useful for testing moderation among treatments

2 Interactive effect interpretation:
   - Does a combination of treatments induce an *additional effect* beyond the sum of separate effects attributable to each treatment?
   - Does being a black female induce an additional effect beyond the effect of being black and that of being female?
   - Useful for finding efficacious treatment combinations in high dimension
An Illustration in the $2 \times 2$ Case

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

$$\begin{align*}
\left[ Y(1, 1) - Y(0, 1) \right] - \left[ Y(1, 0) - Y(0, 0) \right] \\
\text{effect of $A$ when $B = 1$} &\quad \text{effect of $A$ when $B = 0$}
\end{align*}$$

\(\xrightarrow{\sim}\) requires the specification of moderator

- Interactive effect interpretation:

$$\begin{align*}
\left[ Y(1, 1) - Y(0, 0) \right] - \left[ Y(1, 0) - Y(0, 0) \right] - \left[ Y(0, 1) - Y(0, 0) \right] \\
\text{effect of $A$ and $B$} &\quad \text{effect of $A$ when $B = 0$} &\quad \text{effect of $B$ when $A = 0$}
\end{align*}$$

\(\xrightarrow{\sim}\) 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
  - Inference depends on the choice of baseline condition

- $3 \times 3$ 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 $(a_0, b_0)$:
    \[
    \mathbb{E}(Y | A, B) = 1 + a_1^* + a_2^* + b_2^* + a_1^*b_2^* + 2a_2^*b_2^* + 3a_2^*b_1^*
    \]
  - Interaction effect for $(a_2, b_2) >$ Interaction effect for $(a_1, b_2)$
  - Another equivalent model with the baseline condition $(a_0, b_1)$:
    \[
    \mathbb{E}(Y | A, B) = 1 + a_1^* + 4a_2^* + b_2^* + a_1^*b_2^* - a_2^*b_2^* - 3a_2^*b_0^*
    \]
  - 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

<table>
<thead>
<tr>
<th>Job experience</th>
<th>None</th>
<th>4th grade</th>
<th>8th grade</th>
<th>High school</th>
<th>Two-year college</th>
<th>College</th>
<th>Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (baseline)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1–2 years</td>
<td>0</td>
<td>0.009</td>
<td>−0.019</td>
<td>−0.032</td>
<td>0.100</td>
<td>−0.044</td>
<td>−0.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.063)</td>
<td>(0.063)</td>
<td>(0.063)</td>
<td>(0.064)</td>
<td>(0.064)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>3–5 years</td>
<td>0</td>
<td>0.016</td>
<td>0.056</td>
<td>0.165</td>
<td>0.107</td>
<td>0.010</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.063)</td>
<td>(0.064)</td>
<td>(0.064)</td>
<td>(0.064)</td>
<td>(0.065)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>&gt; 5 years</td>
<td>0</td>
<td>−0.050</td>
<td>0.126</td>
<td>0.042</td>
<td>0.058</td>
<td>−0.094</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.064)</td>
<td>(0.064)</td>
<td>(0.063)</td>
<td>(0.064)</td>
<td>(0.064)</td>
<td>(0.064)</td>
</tr>
</tbody>
</table>
The Effects of Changing the Baseline Condition

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

<table>
<thead>
<tr>
<th>Job experience</th>
<th>None</th>
<th>4th grade</th>
<th>8th grade</th>
<th>High school</th>
<th>Two-year college</th>
<th>College</th>
<th>Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.015</td>
<td>0.065</td>
<td>−0.111</td>
<td>−0.027</td>
<td>−0.043</td>
<td>0.109</td>
<td>0</td>
</tr>
<tr>
<td>(0.064)</td>
<td>(0.062)</td>
<td>(0.064)</td>
<td>(0.061)</td>
<td>(0.063)</td>
<td>(0.063)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–2 years</td>
<td>0.078</td>
<td>0.138</td>
<td>−0.066</td>
<td>0.006</td>
<td>0.120</td>
<td>0.129</td>
<td>0</td>
</tr>
<tr>
<td>(0.064)</td>
<td>(0.062)</td>
<td>(0.062)</td>
<td>(0.061)</td>
<td>(0.062)</td>
<td>(0.062)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3–5 years</td>
<td>−0.102</td>
<td>−0.036</td>
<td>−0.172</td>
<td>0.021</td>
<td>−0.054</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>(0.062)</td>
<td>(0.062)</td>
<td>(0.063)</td>
<td>(0.062)</td>
<td>(0.061)</td>
<td>(0.062)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 5 years</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Egami and Imai (Princeton) Causal Interaction AsianPolMeth (Jan., 2016)
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:
  \[ A \in \mathcal{A} = \{a_0, a_1, \ldots, a_{DA-1}\} \]
  \[ B \in \mathcal{B} = \{b_0, b_1, \ldots, b_{DB-1}\} \]

- 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 \(B\)
   \[
   \psi(a_\ell, a_0) \equiv \int_B \mathbb{E}\{Y(a_\ell, B) - Y(a_0, B)\}dF(B)
   \]
   - 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:**
  \[
  \mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_m)\} - \mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\}
  \]
  Effect of \( A = a_\ell \) when \( B = b_m \)
  
  \[
  \mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\}
  \]
  Effect of \( A = a_\ell \) when \( B = b_0 \)

- **Interactive effect interpretation:**
  \[
  \tau(a_\ell, b_m; a_0, b_0) - \mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\} - \mathbb{E}\{Y(a_0, b_m) - Y(a_0, b_0)\}
  \]
  ATCE
  
  \[
  \mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\}
  \]
  Effect of \( A = a_\ell \) when \( B = b_0 \)
  
  \[
  \mathbb{E}\{Y(a_0, b_m) - Y(a_0, b_0)\}
  \]
  Effect of \( B = b_m \) when \( A = a_0 \)

- **Estimation:** Linear regression with interaction terms
The New Causal Interaction Effect

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

\[
\pi(a_\ell, b_m; a_0, b_0) \equiv \tau(a_\ell, b_m; a_0, b_0) - \psi(a_\ell, a_0) - \psi(b_m, b_0)
\]

- 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:

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

- We prove that the **AMTIEs** are both *interval and order invariant*

- The **AMTIEs** 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

<table>
<thead>
<tr>
<th>Job experience</th>
<th>None</th>
<th>4th grade</th>
<th>8th grade</th>
<th>High school</th>
<th>Two-year college</th>
<th>College</th>
<th>Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0</td>
<td>-0.004</td>
<td>-0.028</td>
<td>-0.035</td>
<td>-0.031</td>
<td>0.012</td>
<td>-0.010</td>
</tr>
<tr>
<td>1–2 years</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.025</td>
<td>-0.040</td>
<td>0.024</td>
<td>-0.009</td>
<td>-0.044</td>
</tr>
<tr>
<td>3–5 years</td>
<td>-0.040</td>
<td>-0.019</td>
<td>-0.042</td>
<td>0.031</td>
<td>-0.026</td>
<td>-0.022</td>
<td>0.024</td>
</tr>
<tr>
<td>&gt; 5 years</td>
<td>-0.014</td>
<td>-0.031</td>
<td>0.041</td>
<td>-0.011</td>
<td>-0.021</td>
<td>-0.036</td>
<td>-0.024</td>
</tr>
</tbody>
</table>
AMTIE is Invariant to the Choice of Baseline Condition

<table>
<thead>
<tr>
<th>Job experience</th>
<th>None</th>
<th>4th grade</th>
<th>8th grade</th>
<th>High school</th>
<th>Two-year college</th>
<th>College</th>
<th>Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.024</td>
<td>0.020</td>
<td>−0.004</td>
<td>−0.011</td>
<td>−0.007</td>
<td>0.036</td>
<td>0.014</td>
</tr>
<tr>
<td>1–2 years</td>
<td>0.023</td>
<td>0.023</td>
<td>−0.001</td>
<td>−0.016</td>
<td>0.048</td>
<td>0.015</td>
<td>−0.020</td>
</tr>
<tr>
<td>3–5 years</td>
<td>−0.016</td>
<td>0.005</td>
<td>−0.018</td>
<td>0.055</td>
<td>−0.002</td>
<td>0.002</td>
<td>0.048</td>
</tr>
<tr>
<td>&gt; 5 years</td>
<td>0.010</td>
<td>−0.007</td>
<td>0.065</td>
<td>0.013</td>
<td>0.003</td>
<td>−0.012</td>
<td>0</td>
</tr>
</tbody>
</table>
The Relationships between the ATIE and the AMTIE

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

\[
\pi(a_\ell, b_m; a_0, b_0) = \xi(a_\ell, b_m; a_0, b_0) - \sum_{a \in A} \Pr(A_i = a) \xi(a, b_m; a_0, b_0) \\
- \sum_{b \in B} \Pr(B_i = b) \xi(a_\ell, b; a_0, b_0)
\]

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

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

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

  2. Independent treatment assignment

     $T_j \perp \perp \mathbf{T}_{-j}$ 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
Generalize the 2-way ATIE by marginalizing the other treatments $T^{1:2}$

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

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

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

Higher-order ATIEs are similarly defined sequentially

This representation is based on the conditional effect interpretation

Problem: the conditional effect of conditional effects!
The $K$-way Average Marginal Treatment Interaction Effect

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

\[
\tau_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03}) \]

- \(\text{ATCE}\)

\[
- \left\{ \pi_{1:2}(t_1, t_2; t_{01}, t_{02}) + \pi_{2:3}(t_2, t_3; t_{02}, t_{03}) + \pi_{1:3}(t_1, t_3; t_{01}, t_{03}) \right\} \]

- \(\text{sum of 2-way AMTIEs}\)

\[
- \left\{ \psi(t_1; t_{01}) + \psi(t_2; t_{02}) + \psi(t_3; t_{03}) \right\} \]

- \(\text{sum of (1-way) AMTEs}\)

- **Properties:**
  1. $K$-way ATCE = the sum of all $K$-way and lower-order AMTIEs
  2. Interval and order invariance to the baseline condition
  3. Derive the relationships between the AMTIEs and ATIEs for any order
Empirical Analysis of the Immigration Survey Experiment

- 5 factors (gender$^2$, education$^7$, origin$^{10}$, experience$^4$, plan$^4$)
  - full factorial design assumption
  - 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 $\Rightarrow$ 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 AMTIEs:** 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 $\times$ experience):

$$\tau(\text{Somalia, 1-2 years; India, None}) = \psi(\text{Somalia; India}) + \psi(1-2\text{years; None}) + \pi(\text{Somalia, 1-2years; India, None})$$

$$= -3.74 = -5.14 + 5.12 -3.72$$

- **Three-way examples** (education $\times$ gender $\times$ origin):

$$\tau(\text{Graduate, Male, India; Graduate, Female, India}) = \psi(\text{Male; Female}) + \pi(\text{Graduate, Male; Graduate, Female}) + \pi(\text{Male, India; Female, India}) + \pi(\text{Graduate, Male, India; Graduate, Female, India})$$

$$= 7.46 = -0.77 + -0.34 + 1.56 + 7.01$$
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. $\Rightarrow$ effective analysis of interactions in high-dimension

- Estimation challenges in high dimension
  1. group lasso, hierarchical interaction
  2. post-selection inference


Send comments and suggestions to negami@Princeton.Edu or kimai@Princeton.Edu