Causal Interaction in Factorial Experiments: Application to Conjoint Analysis

Naoki Egami Kosuke Imai

Princeton University

Workshop on the Frontiers of Statistical Analysis

and Formal Theory of Political Sciencethe
Gakusyuin University
January 6, 2017

Causal Heterogeneity and Interaction Effects

- Causal inference revolution in social sciences
  - Randomized experiments: laboratory, field, and survey experiments
  - Observational studies: natural experiments, research designs
- Many methods for estimating average treatment effect (ATE)
- Beyond ATE $\rightsquigarrow$ Causal heterogeneity
  1. Moderation:
     - How does the effect of a treatment vary across individuals?
     - Interaction between the treatment variable and pre-treatment covariates
  2. Causal interaction:
     - What combination of treatments is efficacious?
     - Interaction among multiple treatment variables
  3. Individualized treatment regimes:
     - What treatment combination is optimal for a given individual?

Factorial Experiments for Causal Interaction

- Causal interaction requires multiple treatments
- Randomized experiments with a factorial design
  - Factor = categorical variable with discrete values or "levels"
  - Example: $2^2 \cdot 3 \cdot 4$ design (Gerber and Green, 2000)

	Mail			
	None	Once	Twice	3 times
Phone				
Visit				
Civic	33	103	126	122
Neighbor/civic[a]	74	144	113	127
Close	110	138	113	134
No visit				
Civic	581	443	432	479
Neighbor/civic[a]	0	491	520	542
Close	377	517	534	501
No phone				
Visit				
Civic	1,011	150	213	227
Neighbor	853	175	201	194
Close	822	194	211	206
No visit				
Civic		870	922	825
Neighbor	10,800	764	849	767
Close		722	817	783

- Factorial design is often used for audit studies and conjoint analysis

Conjoint Analysis

- Survey experiments with a factorial design
- Respondents evaluate several pairs of randomly selected profiles defined by multiple factors
- Social scientists use it to analyze multidimensional preferences
- Example: Immigration preference (Hopkins and Hainmueller 2014)
  - representative sample of 1,407 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}
  - What combinations of immigrant characteristics do Americans prefer?
  - High dimension: over 1 million treatment combinations
- Methodological challenges:
  - Many interaction effects $\rightsquigarrow$ false positives, difficulty of interpretation
  - Very few applied researchers study interaction

The Overview of the Talk

1. New causal estimand: Average Marginal Interaction Effect (AMIE)
    - relative magnitude does not depend on baseline condition
    - intuitive interpretation even for high dimension
    - estimation using ANOVA with weighted zero-sum constraints
    - regularization done directly on AMIEs
2. Comparison with the conventional interaction effect:
    - lack of invariance to the choice of baseline condition
    - difficulty of interpretation for higher-order interaction
3. Reanalysis of the conjoint analysis on ethnic voting in Africa

Factorial Experiments with Two Treatments

- Two factorial treatments (e.g., gender and race):

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

- 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

Main Causal Estimands in Factorial Experiments

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

     $$\tau_{AB}(a_\ell, b_m; a_0, b_0) \quad = \quad \mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_0)\}$$

   - Effect of being Asian male

2. Average Marginal Effect (AME; Hainmueller *et al.* 2014; Dasgupta *et al.* 2015):
   - Average effect of treatment $A = a_\ell$ relative to the baseline condition $A = a_0$ averaging over the other treatment B

     $$\psi_A(a_\ell, a_0) \quad = \quad \int \mathbb{E}\{Y(a_\ell, B) - Y(a_0, B)\} dF(B)$$

   - Effect of being male averaging over race

The New Causal Interaction Effect

- Average Marginal Interaction Effect (AMIE):

$$\pi_{AB}(a_\ell, b_m; a_0, b_0) = \underbrace{\tau_{AB}(a_\ell, b_m; a_0, b_0)}_{\text{ACE of } (a_\ell, b_m)} - \underbrace{\psi_A(a_\ell, a_0)}_{\text{AME of } a_\ell} - \underbrace{\psi_B(b_m, b_0)}_{\text{AME of } b_m}$$

- Interpretation: additional effect induced by $A = a_\ell$ and $B = b_m$ together beyond the separate effect of $A = a_\ell$ and that of $B = b_m$
- Additional effect of being Asian male beyond the sum of separate effects for being male and being Asian
- Decomposition of ACE: $\tau_{AB} = \psi_A + \psi_B + \pi_{AB}$
- Invariance: the *relative magnitude* of AMIE does not depend on the choice of baseline condition
- AMIEs depend on the distribution of treatment assignment:
  1. specified by one's experimental design
  2. motivated by a target population

The Conventional Causal Interaction Effect

- Average Interaction Effect (AIE):

$$\xi_{AB}(a_\ell, b_m; a_0, b_0) = \mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_m) - Y(a_\ell, b_0) + Y(a_0, b_0)\}$$

- Equal to linear regression coefficients
- Interactive effect interpretation (similar to AMIE):

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

- Conditional effect interpretation:

$$\begin{aligned} & \mathbb{E}\{Y(a_\ell, b_m) - Y(a_0, b_m)\} - \mathbb{E}\{Y(a_\ell, b_0) - Y(a_0, b_0)\} \\ = \ & \mathbb{E}\{Y(a_\ell, b_m) - Y(a_\ell, b_0)\} - \mathbb{E}\{Y(a_0, b_m) - Y(a_0, b_0)\} \end{aligned}$$

  - difference in effect of being male between Asian and White
  - difference in effect of being Asian between male and female

Comparison between AMIE and AIE

- AIE is NOT invariant to baseline category:
  1. cannot compare regression coefficients
  2. zero interaction when a baseline category is involved

$$\xi_{AB}(a_\ell, b_0; a_0, b_0) \;=\; \xi_{AB}(a_0, b_m; a_0, b_0) \;=\; 0 \quad \text{for all } \ell, m$$

  3. cannot regularize regression coefficients

- AMIE and AIE are closely related:
  1. Conditional effect as a function of AMIE

$$\mathbb{E}\{Y_i(a_\ell, b_0) - Y_i(a_0, b_0)\} \;=\; \psi_A(a_\ell; a_0) + \pi_{AB}(a_\ell, b_0; a_0, b_0)$$

  2. AIE is a linear function of AMIEs

$$\xi_{AB}(a_\ell, b_m; a_0, b_0) \;=\; \pi_{AB}(a_\ell, b_m; a_0, b_0) - \pi_{AB}(a_\ell, b_0; a_0, b_0) - \pi_{AB}(a_0, b_m; a_0, b_0)$$

  3. AMIE is also a linear function of AIEs
  4. No causal interaction $\rightsquigarrow$ zero AMIEs, zero AIEs

Higher-order Causal Interaction

- J factorial treatments with L_j levels each: $\mathbf{T} = (T_1, \ldots, T_J)$
- 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

Higher-order Average Marginal Interaction Effect

- General definition: the difference between ACE and the sum of all lower-order AMIEs (first-order AMIE = AME)
- Example: 3-way AMIE, $\pi_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})$, equals

$$\underbrace{\tau_{1:3}(t_1, t_2, t_3; t_{01}, t_{02}, t_{03})}_{\text{ACE}}$$

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

$$- \underbrace{\left\{\psi(t_1; t_{01}) + \psi(t_2; t_{02}) + \psi(t_3; t_{03})\right\}}_{\text{sum of AMEs}}$$

- Properties:
  1. K-way ACE = the sum of all K-way and lower-order AMIEs
  2. Invariance to the baseline condition

Difficulty of Higher-order AIEs

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

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

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

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

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

Nonparametric Estimation of AMIE

1. Difference-in-means estimator
   - estimate ACE and AMEs using the difference-in-means estimators
   - estimate AMIE as $\hat{\pi}_{AB} = \hat{\tau}_{AB} - \hat{\psi}_A - \hat{\psi}_B$
   - higher-order AMIEs can be estimated sequentially
   - uses the empirical treatment assignment distribution
2. ANOVA based estimator
   - saturated ANOVA include all interactions up to the Jth order
   - weighted zero-sum constraints: for all factors and levels,

$$\sum_{\ell=0}^{L_A-1} \Pr(A_i = a_\ell)\beta_\ell^A = 0, \quad \sum_{\ell=0}^{L_A-1} \Pr(A_i = a_\ell)\beta_{\ell m}^{AB} = 0,$$
$$\sum_{m=0}^{L_B-1} \Pr(B_i = b_m)\beta_m^B = 0, \quad \sum_{m=0}^{L_B-1} \Pr(B_i = b_m)\beta_{\ell m}^{AB} = 0, \quad \text{and so on}$$

   - AMIEs are differences of coefficients:

$$\mathbb{E}(\hat{\beta}_\ell^A - \hat{\beta}_0^A) = \psi_A(a_\ell; a_0), \quad \mathbb{E}(\hat{\beta}_{\ell m}^{AB} - \hat{\beta}_{00}^{AB}) = \pi_{AB}(a_\ell, b_m; a_0, b_0)$$

   - can use any marginal treatment assignment distribution of choice

Regularization via GASH-ANOVA

- Too many coefficients to be estimated $\rightsquigarrow$ over fitting, false positives, difficult interpretation
- Need for regularization by collapsing levels and selecting factors
- Grouping and Selection using Heredity in ANOVA (Post and Bondell):

$$\sum_{\ell,\ell'} w^A_{\ell\ell'} \max\{\phi^A(\ell,\ell')\} + \sum_{m,m'} w^B_{mm'} \max\{\phi^B(m,m')\} \leq \underbrace{c}_{\text{cost parameter}}$$

where

$$\phi^A(\ell,\ell') = |\underbrace{\beta^A_\ell - \beta^A_{\ell'}}_{\text{AME}}| \bigcup \left\{ \bigcup_{m=0}^{L_B-1} |\underbrace{\beta^{AB}_{\ell m} - \beta^{AB}_{\ell' m}}_{\text{AMIE}}| \right\}$$

- The adaptive weight takes the following form:

$$w^A_{\ell\ell'} = \left[(L_A+1)\sqrt{L_A}\max\{\bar{\phi}^A(\ell,\ell')\}\right]^{-1}$$

where $\bar{\phi}^A(\ell,\ell')$ is AMEs and AMIEs estimated without regularization

Conjoint Analysis of Ethnic Voting in Africa

- Ethnic voting and accountability: Carlson (2015, *World Politics*)
- Do voters prefer candidates of same ethnicity regardless of their prior performance? Do ethnicity and performance interact?

- Conjoint analysis in Uganda: 547 voters from 32 villages
- Each voter evaluates 3 pairs of hypothetical candidates
- 5 factors: `Coethnicity`^{2}, `Prior record`^{2}, `Prior office`^{4}, `Platform`^{3}, `Education`^{8}

- `Prior record` = `No` if `Prior office` = `businessman`
  $\rightsquigarrow$ combine these two factors into a single factor with 7 levels
- Collapse `Education` into 2 levels: `relevant degrees` (MA in business, law, economics, development) and `other degrees`

A Statistical Model of Preference Differentials

- ANOVA regression with one-way and two-way effects:

$$Y_i(\mathbf{T}_i) \;=\; \mu + \sum_{j=1}^{J} \sum_{\ell=0}^{L_j-1} \beta_\ell^j \mathbf{1}\{T_{ij} = \ell\} + \sum_{j \neq j'} \sum_{\ell=0}^{L_j-1} \sum_{m=0}^{L_{j'}-1} \beta_{\ell m}^{jj'} \mathbf{1}\{T_{ij} = \ell, T_{ij'} = m\} + \epsilon_i$$

  with appropriate weighted zero-sum constraints
- In conjoint analysis, we observe the sign of preference differentials
- Linear probability model of preference differential:

$$\begin{aligned}
& \Pr(Y_i(\mathbf{T}_i^*) > Y_i(\mathbf{T}_i^\star) \mid \mathbf{T}_i^*, \mathbf{T}_i^\star) \\
= \quad & \mu^* + \sum_{j=1}^{J} \sum_{\ell=0}^{L_j-1} \beta_\ell^j (\mathbf{1}\{T_{ij}^* = \ell\} - \mathbf{1}\{T_{ij}^\star = \ell\}) \\
& + \sum_{j \neq j'} \sum_{\ell=0}^{L_j-1} \sum_{m=0}^{L_{j'}-1} \beta_{\ell m}^{jj'} (\mathbf{1}\{T_{ij}^* = \ell, T_{ij'}^* = m\} - \mathbf{1}\{T_{ij}^\star = \ell, T_{ij'}^\star = m\})
\end{aligned}$$

  where $\mu^* = 0.5$ if the position of profile does not matter
- We apply GASH-ANOVA to this model

Ranges of Estimated AMEs and AMIEs

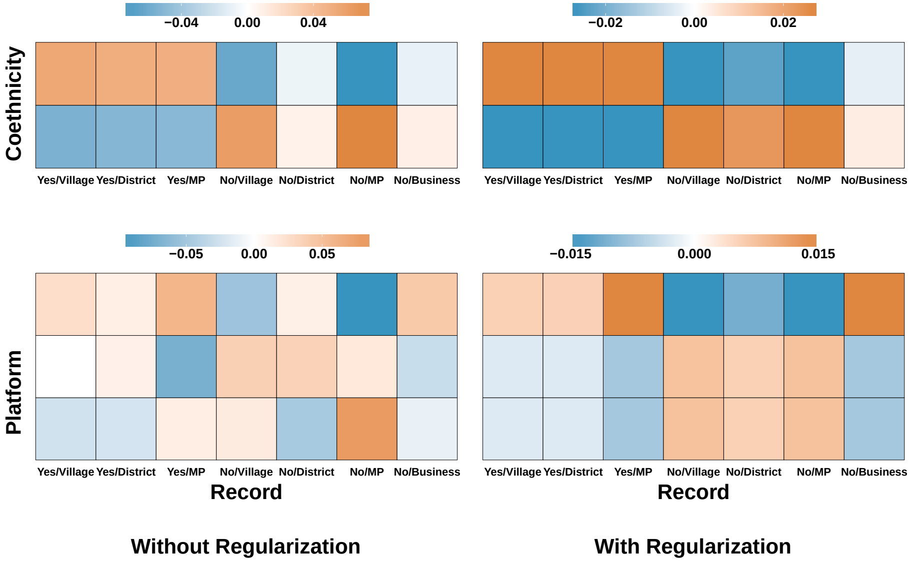
	Range	Selection prob.
AME		
Record	0.122	1.00
Coethnicity	0.053	1.00
Platform	0.023	0.93
Degree	0.000	0.33
AMIE		
Coethnicity × Record	0.053	1.00
Record × Platform	0.030	0.92
Platform × Coethnic	0.008	0.64
Coethnicity × Degree	0.000	0.62
Platform × Degree	0.000	0.35
Record × Degree	0.000	0.09

- Factor selection probability based on bootstrap

Close Look at the Estimated AMEs

Factor	AME	Selection prob.
Record		
Yes/Village	0.122	
		⟩ 0.71
Yes/District	0.122	
		⟩ 0.77
Yes/MP	0.101	
		⟩ 1.00
No/Village	0.047	
		⟩ 0.74
No/District	0.051	
		⟩ 0.74
No/MP	0.047	
		⟩ 1.00
No/Businessman	base	
Platform		
Jobs	−0.023	
		⟩ 0.56
Clinic	−0.023	
		⟩ 0.94
Education	base	
Coethnicity	0.054	1.00
Degree	0.000	0.33

Effect of Regularization on AMIEs



Decomposition and Conditional Effects

- Decomposition of ACE (Coethnicity × Record interaction):

$$\underbrace{\tau(\texttt{Coethnic, No/Business}; \texttt{Non-coethnic, No/MP})}_{-2.4}$$

$$= \underbrace{\psi(\texttt{Coethnic}; \texttt{Non-coethnic})}_{5.4} + \underbrace{\psi(\texttt{No/Business}; \texttt{No/MP})}_{-4.7}$$

$$+ \underbrace{\pi(\texttt{Coethnic, No/Business}; \texttt{Non-coethnic, No/MP})}_{-3.1}$$

- Conditional effects (Platform × Record interaction):
  - AMIE: $\pi(\texttt{Education, No/MP}\}; \{\texttt{Job, No/MP}\}) = -2.3$
  - Conditional effect of Education relative to Job for No/MP is approximately zero
  - AME: $\psi(\texttt{Education}; \texttt{Job}) = 2.3$

Concluding Remarks

- Interaction effects play an essential role in causal heterogeneity
  1. moderation
  2. causal interaction
- Randomized experiments with a factorial design
  1. useful for testing multiple treatments and their interactions
  2. social science applications: audit studies, conjoint analysis
  3. challenge: estimation and interpretation in high dimension
- Average Marginal Interaction Effect (AMIE)
  1. invariant to baseline condition
  2. straightforward interpretation even for high order interaction
  3. enables effect decomposition
  4. enables regularization through ANOVA
- Designing factorial experiments (work in progress)
  1. select factors and levels via our method to reduce dimension
  2. use unregularized ANOVA for the main study

References

1. Egami, Naoki and Kosuke Imai. "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis." Working paper available at http://imai.princeton.edu/research/int.html

2. Egami, Naoki, Marc Ratkovic, and Kosuke Imai. "FindIt: Finding Heterogeneous Treatment Effects." R package available at CRAN

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