The Identification and Estimation of Direct and Indirect Effects in A/B Tests through Causal Mediation Analysis

Xuan Yin [1]   Liangjie Hong [2]

[1]xyin@etsy.com

[2]lhong@etsy.com

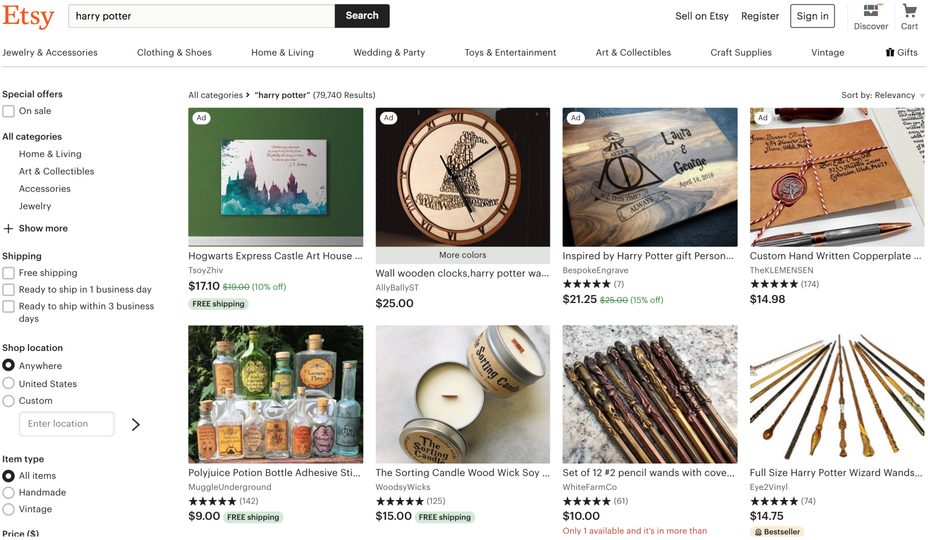
August 14, 2019

Xuan Yin and Liangjie Hong. 2019. The Identification and Estimation of Direct and Indirect Effects in A/B Tests through Causal Mediation Analysis. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (KDD '19). ACM, New York, NY, USA, 2989-2999. DOI: `https://doi.org/10.1145/3292500.3330769`

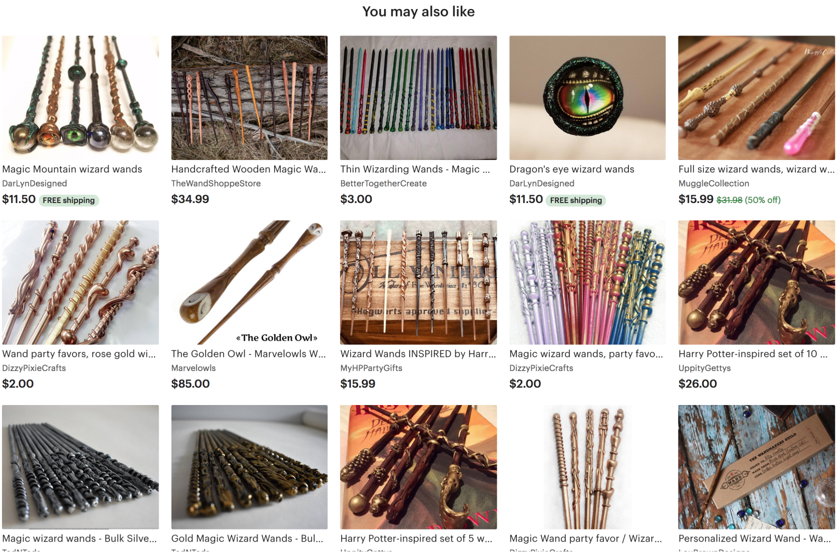
Overview: What is the research about?

- **Background**: User engagement of different products can be causally dependent.
- **Goal**: To propose new KPI that takes care of the causal dependency.
- **How**: Through the causal identification and estimation of direct and indirect effects using data of A/B tests

Introduction: Examples of Online Products: Organic Search and Promoted Listings



Introduction: Examples of Online Products: Recommendation Module



Introduction
We see causal dependency from A/B test results.

- ***Induced Change***: A change in one product would ***induce*** users to change their behaviors in other products.
- Examples I:

Table: **Recommendation Module** A/B Test Average Treatment Effect (**ATE**)

Number of clicks on **Recommendation Modules**	Signficant ↑
Number of clicks on **Organic Search** results	Significant ↓
Conversion/Gross merchandise value (GMV)	**Insignificant** Change

Introduction
We see causal dependency from A/B test results.

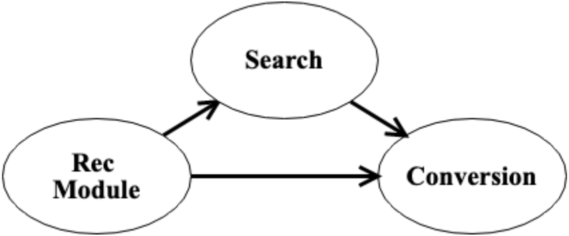
- Examples II:

Table: **Promoted Listing** A/B Test Average Treatment Effect (**ATE**)

Promoted Listing	
click-through-rate	Significant ↑
number of clicks	Significant ↑
advertising revenue	Significant ↑
Number of clicks on **Organic Search** results	Significant ↓
Conversion/Gross merchandise value (GMV)	**Insignificant** Change

Introduction: The most popular KPI is **ATE** from A/B tests

Suppose the underlying causal mechanism is like



Questions:

- Does **ATE on Conversion** truly measure the contribution of rec module change to the marketplace?
- Is **ATE on Conversion** still a good KPI for rec module?
- Shall we just ignore the induced reduction in user engagement of search?

Introduction: Problems of Funnel Analysis

Many e-commerce companies use **Tight Attribution Metric as KPI**.

- purchase funnel: click A in rec module$\Rightarrow$purchase A

Problems: Too Heuristic, No Foundation

- **Ambiguous**
  click A in rec module
  $\Rightarrow$click A in search results
  $\Rightarrow$click A in many different places
  $\Rightarrow$purchase A
  Which place shall get the point?
- **Too Narrow**
  view the rec module, dwell time $\uparrow$, but never click it
  $\Rightarrow$purchase sth elsewhere
  Shall rec module get any point?

Introduction: Problems of Funnel Analysis

The severest problem of funnel analysis in A/B tests:

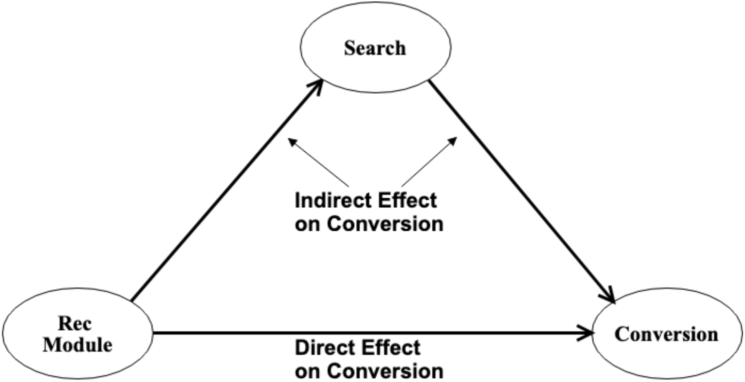
It may destroy the causal interpretation of experimental results.

Because

- It subsets the experimental results based on post-treatment criteria.
- Conditional on post-treatment variable, the randomization of treatment assignment may no longer hold.
  (i.e., it could break **ignorability** of the identification of **ATE**)
  See, e.g., Montgomery et al. (2018)

Introduction: Direct and Indirect Effects

How about we split **ATE** to two parts: **Direct Effect** and **Indirect Effect**?



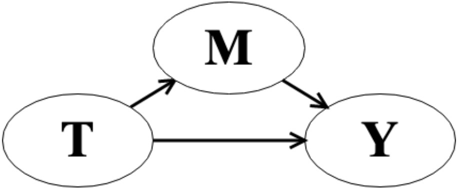
Use **Direct Effect** on Conversion as KPI!

Introduction

- A/B tests cannot give us **Direct Effect** or **Indirect Effect**.
- It can only identify **ATE**.
- To conduct analysis, We need to formalize the idea using formal causal inference language.

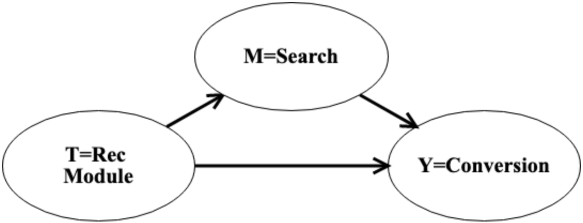
Introduction to Potential Outcome Framework

- In an A/B test, a user i is randomly assigned to either treatment group ($T_i = 1$) or control group ($T_i = 0$).
- Let $M_i(t)$ denote her potential mediator under treatment t.
- Let $Y_i(t, m)$ denote her potential outcome under the treatment t and the mediator m.
- Only one of potential mediators and only one of potential outcomes can be observed for each user.



Examples of Potential Outcomes

In recommendation module A/B tests,



- $M_i(1)$ is her numbers of clicks on search results if she was presented with the new recommendation module.
- $Y_i(1, M_i(0))$ is her conversion status if she was presented with the new rec module and clicked on search results as if she had been presented with the old one.

Causal Identification

The Fundamental Research Question of Causal Inference Is **Identification**.

- Causal effect: the difference between potential outcomes.

Causal Identification

Assumptions $\Rightarrow$ **Causal Effects**

Example: Identification in Rubin Causal Model
The Model Behind A/B Tests

Identification of ATE

Strong Ignorability **and** ***SUTVA*** $\Rightarrow$ **ATE**

- **ATE** on $Y := \mathbb{E}(Y_i(1, M_i(1))) - \mathbb{E}(Y_i(0, M_i(0)))$
- **ATE** on $M := \mathbb{E}(M_i(1)) - \mathbb{E}(M_i(0))$
- ***Strong Ignorability***: $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i$ and $0 < \mathbb{P}(T_i = t) < 1$
- ***SUTVA***: *Stable Unit-Treatment-Value Assumption*

Causal Mediation Analysis (**CMA**)

Average Direct Effect (ADE)

$$\mathbf{ADE}(t) := \mathbb{E}(Y_i(1, M_i(t))) - \mathbb{E}(Y_i(0, M_i(t)))$$

- $\mathbf{ADE}(0)$ is the **direct effect** of the rec module change on conversion **leaving aside the induced change**.
- Because mediator is fixed at $M(t)$, the difference between the two potential outcomes can only be attributed to the two different treatments.

Causal Mediation Analysis (**CMA**)

Average Causal Mediation Effect (ACME, Indirect Effect)

$$\textbf{ACME}(t) := \mathbb{E}(Y_i(t, M_i(1))) - \mathbb{E}(Y_i(t, M_i(0)))$$

- $\textbf{ACME}(1)$ is the average effect of the *induced change* in organic search clicks upon conversion given users were presented with the new rec module all the time.
- Because treatment is fixed at t, the difference between the two potential outcomes can only be attributed to the two different potential mediators, which are *induced* by different treatments.

Causal Mediation Analysis (**CMA**)

Identification of Direct and Indirect Effects in **CMA** (Imai et al., 2010)

***Sequential Ignorability (SI)* and *SUTVA* $\Rightarrow$ ACME and ADE**

SI: add two extra conditions to ***Strong Ignorability***:

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t$$
$$0 < \mathbb{P}(M_i(t) = m | T_i = t) < 1$$

Conditional on the treatment, each potential mediator behaves like the treatment and is ignorable to any potential outcomes.

We Cannot Use **CMA** Directly in A/B Tests

- Multiple unmeasured causally-dependent mediators in A/B tests break ***SI*** and invalidates **CMA**.
- **Fat Hand** (Peysakhovich and Eckles, 2018)

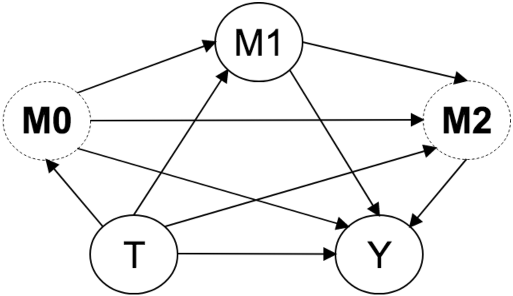


Figure: **M0** and **M2** are unmeasured upstream and downstream mediators of M1

What we do

- The literature of **CMA** is only a starting point.
- We propose new measures for direct and indirect effects.
- We work out the assumptions that lead to new measures.
- We do the estimation and hypothesis testing using real data.
- We prove that

Generalize CMA

Generalized SI and ***LSEM*** $\Rightarrow$ **GADE** and **GACME**

What we do: New Direct Effect

Generalized Average Direct Effect (GADE)

$$\begin{aligned}\textbf{GADE}(t) &= \mathbb{E}[Y_i(1, \boldsymbol{M}_{i0}(1), M_{i1}(t, \boldsymbol{M}_{i0}(t)), \\ &\qquad \boldsymbol{M}_{i2}(1, \boldsymbol{M}_{i0}(1), M_{i1}(t, \boldsymbol{M}_{i0}(t))))] \\ &\quad - \mathbb{E}[Y_i(0, \boldsymbol{M}_{i0}(0), M_{i1}(t, \boldsymbol{M}_{i0}(t)), \\ &\qquad \boldsymbol{M}_{i2}(0, \boldsymbol{M}_{i0}(0), M_{i1}(t, \boldsymbol{M}_{i0}(t))))]\end{aligned}$$

- It captures the causal effect of the treatment T_i that goes through all the channels that do not have M_{i1}:
  $T \to Y$
  $T \to \boldsymbol{M}_0 \to Y$
  $T \to \boldsymbol{M}_0 \to \boldsymbol{M}_2 \to Y$
  $T \to \boldsymbol{M}_2 \to Y$

What we do: New Indirect Effect

Generalized Average Causal Mediation Effect (GACME, Indirect Effect)

$$\begin{aligned}\textbf{GACME}(t) = \; & \mathbb{E}[Y_i(t, \boldsymbol{M}_{i0}(t), M_{i1}(1, \boldsymbol{M}_{i0}(1)), \\ & \qquad \boldsymbol{M}_{i2}(t, \boldsymbol{M}_{i0}(t), M_{i1}(1, \boldsymbol{M}_{i0}(1))))] \\ & - \mathbb{E}[Y_i(t, \boldsymbol{M}_{i0}(t), M_{i1}(0, \boldsymbol{M}_{i0}(0)), \\ & \qquad \boldsymbol{M}_{i2}(t, \boldsymbol{M}_{i0}(t), M_{i1}(0, \boldsymbol{M}_{i0}(0))))]\end{aligned}$$

- It captures the causal effect of the treatment T_i that goes through all the channels that have M_{i1}:
  $T \to M_1 \to Y$
  $T \to M_1 \to \boldsymbol{M}_2 \to Y$
  $T \to \boldsymbol{M}_0 \to M_1 \to Y$
  $T \to \boldsymbol{M}_0 \to M_1 \to \boldsymbol{M}_2 \to Y.$

What we do: The Identification Assumptions

- ***Generalized SI***: Each potential mediator, conditional on the treatment and its upstream mediators, behave like the treatment and is ignorable to all the potential outcomes and all the potential downstream mediators.
- ***LSEM***: *Linear Structural Equation Model*. Potential mediators, potential outcomes, and treatment have linear relationships.

Generalized SI and ***LSEM*** $\Rightarrow$ **GADE** and **GACME**

What we do: How to Estimate Using Real Data

- We estimate **GACME** and **GADE** by General Method of Moments.

Definition (Estimation via Two Linear Regression Equations)

$$\begin{aligned} M_{i1} &= \theta_{M_1 0} + \theta_{M_1 1} T_i + \mu_{M_1} \\ Y_i &= \theta_{Y0} + \theta_{Y1} T_i + \theta_{Y2} M_{i1} + \theta_{Y3} M_{i1} T_i + \mu_Y \\ \mathbf{GADE}(t) &= \theta_{Y1} + \theta_{Y3}(\theta_{M_1 0} + \theta_{M_1 1} t) \\ \mathbf{GACME}(t) &= \theta_{M_1 1}(\theta_{Y2} + \theta_{Y3} t) \end{aligned}$$

- Easy to implement in practice (just two linear regression equations!)
- Just use the data from existing A/B test
- No requirements on extra randomization/intervention

What we do: How to do Hypothesis Testing

- We estimate the asymptotic variances of estimators by Delta method.
- We test H_0: **GADE** $= 0$ and H_0: **GACME** $= 0$ based on asymptotic normality.

What we do: The Relationship to The Literature

Case 1: No Unmeasured Upstream and Downstream Mediators

GADE and **GACME** collapse to **ADE** and **ACME**.

Case 2: Unmeasured Upstream or Downstream Mediator

We cannot identify **ADE** and **ACME**.
However, we can identify **GADE** and **GACME**

In practice, difficult and costly to know or to estimate the upstream or downstream mediators

Estimates of Causal Effects for Recommendation Module A/B Test Mediator is Organic Search Clicks

	Outcome: Conversion	
Effect	% Change	Std Error
GADE(0)	0.4959%*	0.000272
GADE(1)	0.4905%*	0.000271
GACME(0)	-0.2703%***	0.000047
GACME(1)	-0.2757%***	0.000049
ATE	0.2202%	0.000275

1) % Change = Effect/Mean of Control
2) '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$, '.' $p < 0.1$. Two-tailed p-value is derived from z-test for H_0: the effect is zero, which is based on asymptotical normality.

Estimates of Causal Effects for Promoted Listing A/B Test
Mediator is Organic Search Clicks

	Outcome: Conversion	
Effect	% Change	Std Error
GADE(0)	-0.1448%	0.000203
GADE(1)	-0.1472%	0.000202
GACME(0)	-0.2237%***	0.000034
GACME(1)	-0.2261%***	0.000034
ATE	-0.3709%	0.000205

1) % Change = Effect/Mean of Control
2) '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$, '.' $p < 0.1$. Two-tailed p-value is derived from z-test for H_0: the effect is zero, which is based on asymptotical normality.

Take-Aways

- User engagement of different products can be causally dependent.
- The current popular KPI in A/B tests: **ATE** (on Conversion) is undesirable to evaluate product change.
- Tight attribution metric from funnel analysis is not causally interpretable.
- **Direct** and **indirect effects** from **CMA** are desirable, but cannot be identified b/c **fat hand** of A/B tests.
- **GADE** and **GACME** are better KPI for evaluation purposes.
- They can be identified and easily estimated and tested in practice.

Xuan Yin and Liangjie Hong. 2019. The Identification and Estimation of Direct and Indirect Effects in A/B Tests through Causal Mediation Analysis. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (KDD '19). ACM, New York, NY, USA, 2989-2999. DOI: `https://doi.org/10.1145/3292500.3330769`

Imai, K., L. Keele, and T. Yamamoto (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*.

Montgomery, J. M., B. Nyhan, and M. Torres (2018). How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It. *American Journal of Political Science 62*(3), 760–775.

Peysakhovich, A. and D. Eckles (2018). Learning causal effects from many randomized experiments using regularized instrumental variables. In *The Web Conference 2018 (WWW 2018)*, New York, NY. ACM.