Estimation of Monotone Effects in Network Experiments

David Choi

Carnegie Mellon University

June 19, 2017

Experiments with Interference

- Goal: learn the effect of a treatment on a social outcome
- Social means that participants interact with each other. Outcomes are not independent!
  - Sometimes called **interference**
- Examples
  - Education – “peer effects”
  - Health (vaccination trials) – “herd immunity”
  - Advertising – “viral marketing”
  - Facebook experiments on user experience

The Main Challenge

Challenge:

- How do we form statistically significant conclusions from dependent observations?
- Without unreasonable assumptions on the dependence model?

Our Approach

- Assume effects are **monotone** – “treatments never hurt”
  - Either directly or by spillovers
- Allow the interference to be **arbitrary** in all other respects
  - long range, nonlinear, etc.
- Find **one-sided confidence interval** on a **particular** counterfactual of interest
  - “if (no, all) units were treated, what would the outcome be?”
- To improve estimates (i.e., detect spillovers), use **any prior knowledge** to choose the **test statistic**.
  - Safer than making prior assumptions

Formulation

Notation:

- X_i: treatment of ith unit (binary)
- Y_i: outcome of ith unit (binary)
- θ_i: what would have happened to ith unit, in the absence of all treatments (i.e, if $X_i = 0$ for all i)?

Assumptions:

- X is random (sampled w/o replacement)
- $\theta_i \leq Y_i$ for all i ("treatments never hurt")

Note: **Not** assuming anything about who influences whom

Formulation

Notation:

- X_i: treatment of ith unit (binary)

$$X_i = \begin{cases} 1 & i \text{ is treated} \\ 0 & i \text{ is not treated} \end{cases}$$

- Y_i: outcome of ith unit (binary)
- θ_i: what would have happened to ith unit, in the absence of all treatments (i.e, if $X_i = 0$ for all i)?

Assumptions:

- X is random (sampled w/o replacement)
- $\theta_i \leq Y_i$ for all i ("treatments never hurt")

Note: **Not** assuming anything about who influences whom

Formulation

Notation:

- X_i: treatment of ith unit (binary)
- Y_i: outcome of ith unit (binary)

$$Y_i = \begin{cases} 1 & i \text{ has positive outcome} \\ 0 & i \text{ has negative outcome} \end{cases}$$

- θ_i: what would have happened to ith unit, in the absence of all treatments (i.e, if $X_i = 0$ for all i)?

Assumptions:

- X is random (sampled w/o replacement)
- $\theta_i \leq Y_i$ for all i ("treatments never hurt")

Note: **Not** assuming anything about who influences whom

Formulation

Notation:

- X_i: treatment of ith unit (binary)
- Y_i: outcome of ith unit (binary)
- θ_i: what would have happened to ith unit, in the absence of all treatments (i.e, if $X_i = 0$ for all i)?

$$\theta_i = \begin{cases} 1 & i \text{ has positive outcome under counterfactual} \\ 0 & \text{otherwise} \end{cases}$$

Assumptions:

- X is random (sampled w/o replacement)
- $\theta_i \leq Y_i$ for all i ("treatments never hurt")

Note: **Not** assuming anything about who influences whom

Formulation

Notation:

- X_i: treatment of ith unit (binary)
- Y_i: outcome of ith unit (binary)
- θ_i: what would have happened to ith unit, in the absence of all treatments (i.e, if $X_i = 0$ for all i)?

Assumptions:

- X is random (sampled w/o replacement)
- $\theta_i \leq Y_i$ for all i ("treatments never hurt")

Note: **Not** assuming anything about who influences whom

Two Methods

In paper, we propose two methods for constructing a confidence interval in this setting.

1. **Inverting a test statistic**
2. Normal-based confidence intervals

(Will only have time to present first method)

Note: At this point, the goal is to show proof of concept, as opposed to a solution that works “out of the box” for applications.

[simple idea that doesn't work]

Spoiler Alert

Story will be similar for both methods:

- Without network information, one-sided CIs usually similar to assuming SUTVA, but with provable coverage
- With "good" network information, CIs can be **tightened** through choice of test statistic
  - Without placing formal assumptions on the generative model
  - Coverage is preserved, even if network information is only crude proxy to true social mechanisms – or even arbitrarily misspecified

Inverting a Test Statistic

- Random vector X
- Unknown parameter vector θ (the counterfactual)
- Test statistic $T(X;\theta)$, with 95% quantile $t_{.95}(\theta)$
- Null hypothesis θ_{null}

We can reject θ_{null} with 95% confidence if:

$$T(X;\theta_{\text{null}}) > t_{.95}(\theta_{\text{null}}) \qquad \text{or} \qquad \theta_{\text{null}} \not\leq Y$$

A 95% confidence set for θ is the set of all non-rejected hypotheses:

$$\left\{\theta : T(X;\theta) \leq t_{.95}(\theta) \text{ and } \theta \leq Y\right\}$$

This is valid for **any** choice of T (but may be hard to compute)

Inverting a Test Statistic

- Random vector X
- Unknown parameter vector θ (the counterfactual)
- Test statistic $T(X;\theta)$, with 95% quantile $t_{.95}(\theta)$
- Null hypothesis θ_{null}

We can reject θ_{null} with 95% confidence if:

$$T(X;\theta_{\text{null}}) > t_{.95}(\theta_{\text{null}}) \qquad \text{or} \qquad \theta_{\text{null}} \not\leq Y$$

A 95% confidence set for θ is the set of all non-rejected hypotheses:

$$\left\{\theta : T(X;\theta) \leq t_{.95}(\theta) \text{ and } \theta \leq Y\right\}$$

This is valid for **any** choice of T (but may be hard to compute)

Inverting a Test Statistic

- Random vector X
- Unknown parameter vector θ (the counterfactual)
- Test statistic $T(X;\theta)$, with 95% quantile $t_{.95}(\theta)$
- Null hypothesis θ_{null}

We can reject θ_{null} with 95% confidence if:

$$T(X;\theta_{\mathsf{null}}) > t_{.95}(\theta_{\mathsf{null}}) \qquad \text{or} \qquad \theta_{\mathsf{null}} \not\leq Y$$

A 95% confidence set for θ is the set of all non-rejected hypotheses:

$$\left\{\theta : T(X;\theta) \leq t_{.95}(\theta) \text{ and } \theta \leq Y\right\}$$

This is valid for **any** choice of T (but may be hard to compute)

Formulation as Optimization Problem

To find the upper bound of our confidence set on $\sum_i \theta_i$, we can solve the following optimization problem:

$$\begin{aligned} \max_{\theta \in \{0,1\}^N} \quad & \sum_i \theta_i \\ \text{such that} \quad & T(X;\theta) \leq t_{.95}(\theta) \\ & \theta_i \leq Y_i \quad \text{for all } i \end{aligned}$$

The difference $\sum_i Y_i - \sum_i \theta_i$ is a lower bound on the attributable treatment effect[1].

[1] Rosenbaum, *Biometrika* 2001

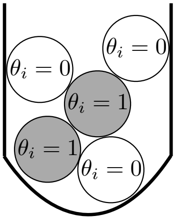
T_{basic}: a Basic Test Statistic

- Let T_{basic} denote a test statistic:

$$T_{\text{basic}}(X;\theta) = \sum_{i:X_i=1} \theta_i$$

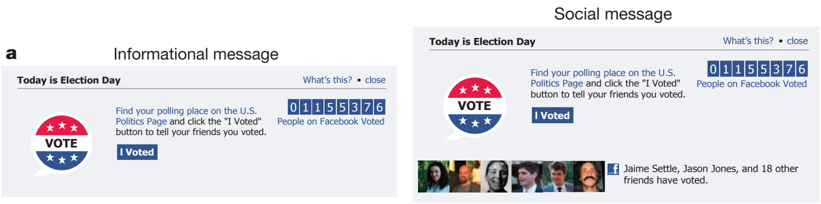
Interpretation: how many treated people have $\theta_i = 1$?

- Distibution of T_{basic} is **hypergeometric**: how many balls with $\theta_i = 1$ are drawn from an urn?



- Hence, a hypergeometric test is a valid test for any θ_{null}

Facebook 2010 Election Experiment[2]



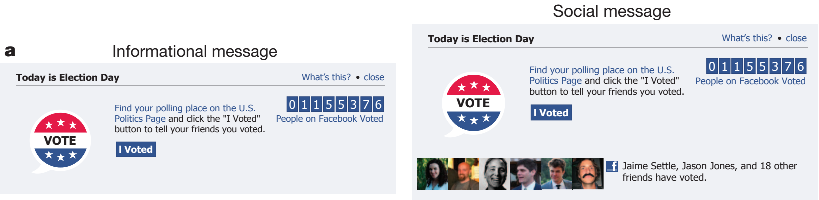
- On login, Facebook users shown advert with "I voted" button
- For some users, the advertisement included profile pictures of friends who had already clicked the button
- Did this make them more likely to do so themselves?

Sources of interference:

- Content of advertisement depends on actions of previous recipients
- Traditional word-of-mouth

[2] Bond. et. al, *Nature*, 2012.

Facebook 2010 Election Experiment[2]



- On login, Facebook users shown advert with "I voted" button
- For some users, the advertisement included profile pictures of friends who had already clicked the button
- Did this make them more likely to do so themselves?

Sources of interference:

- Content of advertisement depends on actions of previous recipients
- Traditional word-of-mouth

[2] Bond. et. al, *Nature*, 2012.

Analysis

- $X_i = 1$ if person i saw profile pictures of friends who voted
- $Y_i = 1$ if person i clicked "I voted" button
- $\theta_i = 1$ if i would have clicked button under full control

	$X_i = 0$	$X_i = 1$
Total	611 K	60 M
$Y_i = 1$	109 K	12 M
Hypothesized $\theta_i = 1$ percentage		

Analysis: find the non-rejected values for θ

Analysis

- $X_i = 1$ if person i saw profile pictures of friends who voted
- $Y_i = 1$ if person i clicked "I voted" button
- $\theta_i = 1$ if i would have clicked button under full control

	$X_i = 0$	$X_i = 1$
Total	611 K	60 M
$Y_i = 1$	109 K	12 M
Hypothesized $\theta_i = 1$	109 K	12 M
percentage	17.8%	20%

reject: p-val $= 0$

Analysis

- $X_i = 1$ if person i saw profile pictures of friends who voted
- $Y_i = 1$ if person i clicked "I voted" button
- $\theta_i = 1$ if i would have clicked button under full control

	$X_i = 0$	$X_i = 1$
Total	611 K	60 M
$Y_i = 1$	109 K	12 M
Hypothesized $\theta_i = 1$	109 K	10.8 M
percentage	17.8%	18%

don't reject: p-val = 0.05

Analysis

- $X_i = 1$ if person i saw profile pictures of friends who voted
- $Y_i = 1$ if person i clicked "I voted" button
- $\theta_i = 1$ if i would have clicked button under full control

	$X_i = 0$	$X_i = 1$
Total	611 K	60 M
$Y_i = 1$	109 K	12 M
Hypothesized $\theta_i = 1$	122 K	12 M
percentage	20%	20%

reject: $\theta \not\preceq Y$

Analysis

- $X_i = 1$ if person i saw profile pictures of friends who voted
- $Y_i = 1$ if person i clicked "I voted" button
- $\theta_i = 1$ if i would have clicked button under full control

	$X_i = 0$	$X_i = 1$
Total	611 K	60 M
$Y_i = 1$	109 K	12 M
Hypothesized $\theta_i = 1$	109 K	10.8 M
percentage	17.8%	18%

don't reject: p-val $= 0.05$

After trying all possible choices, this was the largest non-rejected value of $\sum \theta_i$

Thus, assuming $\theta \leq Y$ yields one-sided CI: $\sum(Y_i - \theta_i) \geq 1.2\,\mathrm{M}$

[more]

Limitations of T_{basic}

$$T_{\text{basic}}(X; \theta) = \sum_{i:\text{treated}} \theta_i$$

T_{basic} does not use any spatial or network information

- As a result, it can only count direct effects
- No power to detect spillovers
- Cannot rule out the possibility of no interference, so confidence interval must include it
  - [1.2M, 1.3M] for this example

Next: new test statistic T_{spill} that uses network information

T_{spill}: a Statistic to Detect Spillovers

- Suppose we have geographic or network data.
- Let T_{spill} equal

“How many people that were near a treated unit would have still had the outcome in the absence of all treatments”

$$T_{\text{spill}}(X;\theta) = \sum_i \sum_j X_i \theta_j \cdot f(dist(i,j)),$$

where

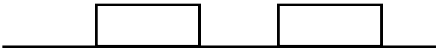
- $f \geq 0$ is a kernel function
- $dist(i,j)$ is the geographic or network distance between i and j

Picture of T_{spill}

- **Task:** search over all $\theta \leq Y$ for non-rejected hypotheses

Treatment	X	0 0 0 0 0 0 **1** 0 0 0 0 0 0 0 0 **1** 0 0 0 0
Outcome	Y	0 **1** 0 0 **1 1 1 1 1** 0 0 0 0 **1 1 1 1 1** 0 0
Counterfactual	θ	0 **?** 0 0 **? ? ? ? ?** 0 0 0 0 **? ? ? ? ?** 0 0

f(dist(i, j))
$= 1\{\text{dist} \leq 3\}$



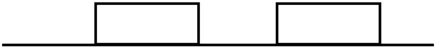
Picture of T_{spill}

- **Task:** search over all $\theta \leq Y$ for non-rejected hypotheses

Treatment	X	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	**1**	0	0	0	0
Outcome	Y	0	**1**	0	0	**1**	**1**	**1**	**1**	**1**	0	0	0	0	**1**	**1**	**1**	**1**	**1**	0	0
Counterfactual	θ	0	**1**	0	0	**1**	**1**	**1**	**1**	**1**	0	0	0	0	**1**	**1**	**1**	**1**	**1**	0	0

f(dist(i, j))

$= 1\{\text{dist} \leq 3\}$



$$T_{\text{spill}}(X; \theta) = 10$$

$$\text{p-val} = 0.005$$

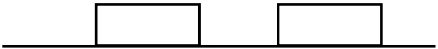
Picture of T_{spill}

- **Task:** search over all $\theta \leq Y$ for non-rejected hypotheses

Treatment	X	0 0 0 0 0 0 **1** 0 0 0 0 0 0 0 0 **1** 0 0 0 0
Outcome	Y	0 **1** 0 0 **1 1 1 1 1** 0 0 0 0 **1 1 1 1 1** 0 0
Counterfactual	θ	0 **1** 0 0 **0 1 1 1 1** 0 0 0 0 **1 1 1 1 1** 0 0

f(dist(i, j))

$= 1\{\text{dist} \leq 3\}$



$$T_{\text{spill}}(X; \theta) = 9$$

$$\text{p-val} = 0.02$$

Picture of T_{spill}

- **Task:** search over all $\theta \leq Y$ for non-rejected hypotheses

Treatment	X	0 0 0 0 0 0 **1** 0 0 0 0 0 0 0 0 **1** 0 0 0 0
Outcome	Y	0 **1** 0 0 **1 1 1 1 1** 0 0 0 0 **1 1 1 1 1** 0 0
Counterfactual	θ	0 **1** 0 0 **0 1 1 1 1** 0 0 0 0 **1 1 1 1 1** 0 0

f(dist(i, j))
$= 1\{\text{dist} \leq 3\}$

$$T_{\text{spill}}(X;\theta) = 9$$

$$\text{p-val} = 0.02$$

- Conceptually, we could check every possible value for θ this way. In practice, this is computationally hard and we'll require an approximation [algorithm]

Pros and Cons of T_{spill}

Good

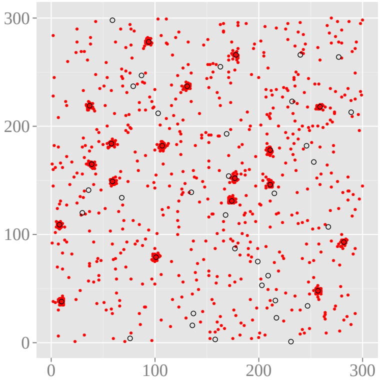
- T_{spill} can detect spillovers
- No exposure model is assumed
  - CI is never anti-conservative as long as effects are monotone

Bad

- CI can be vacuously large, if
  - Kernels are too small or too large (so prior knowledge needed)
  - Computational approximation is too conservative

Simulations

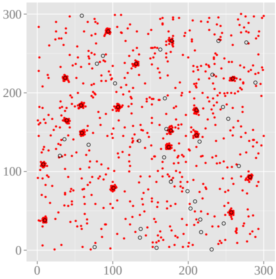
Example simulation:



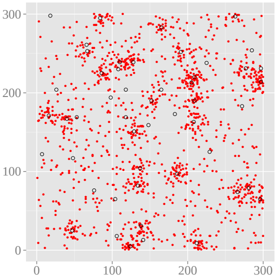
Units live on 300 x 300 grid. Black circles are treatments. Red dots are positive outcomes. Attributable treatment effect $A \approx 600$.

Three examples

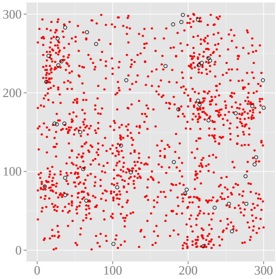
Varying σ_h, the radius of treatment effect



Highly localized effects
$\sigma_h = 3$



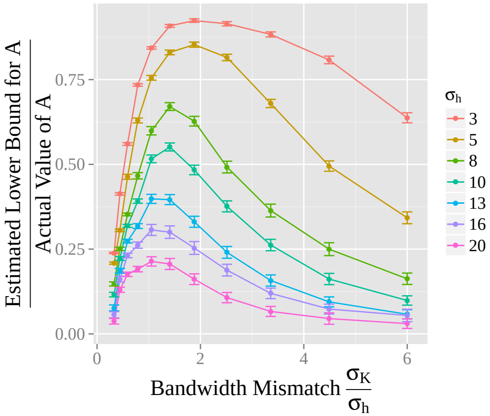
Medium localization
$\sigma_h = 10$



Diffuse effects
$\sigma_h = 20$

Which simulation is easiest for T_{spill}? What if σ_h is misspecified?

Results

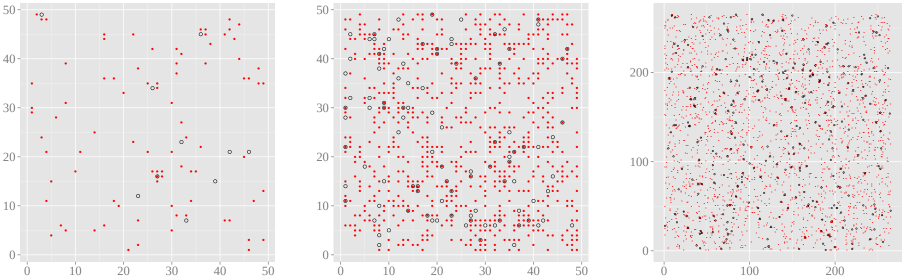


σ_h: actual bandwidth. σ_K: value used in T_{spill}

Localized effects are much easier to estimate than diffuse ones

Three more examples

Varying the number of treatments L and units N



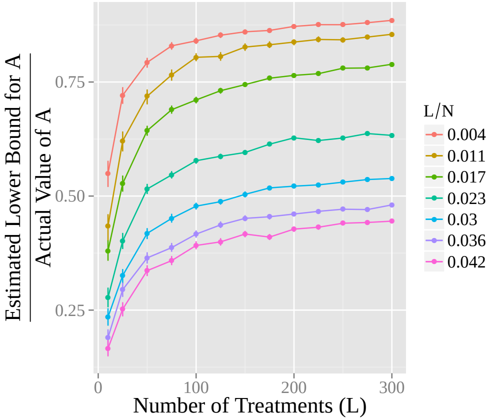
Small L and sparse
$L = 10, \frac{L}{N} = 0.004$

Large L and dense
$L = 100, \frac{L}{N} = 0.04$

Large L and sparse
$L = 300, \frac{L}{N} = 0.004$

Which simulation is easiest for T_{spill}?

Results



Good results when treatments cause many well-separated clusters of outcomes. Infill asymptotic performance can be bad (control group is lost)

Outline

Goal of T_{spill}: proof of concept that prior knowledge can be used to select test statistic, instead of assuming a generative model.

1. Inverting a test statistic
2. **Normal-based confidence intervals**

Normal-based methods: developed for a particular dataset where T_{spill} was bad.

(probably stop here due to time constraints)

Recap

Without spatial or network information, CI includes range given by methods that assume SUTVA:

- Necessary since SUTVA cannot be ruled out
- Any difference in upper bounds could be either because
  - SUTVA might be anti-conservative due to interference, **or**
  - new method might be conservative

With such information, new methods can give improved estimates that rule out hypothesis of no interference

- Without placing formal assumptions on the generative model
- Confidence intervals will have correct coverage, even if network information is only crude proxy to true social mechanisms

Recap

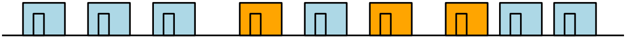
Without spatial or network information, CI includes range given by methods that assume SUTVA:

- Necessary since SUTVA cannot be ruled out
- Any difference in upper bounds could be either because
  - SUTVA might be anti-conservative due to interference, **or**
  - new method might be conservative

With such information, new methods can give improved estimates that rule out hypothesis of no interference

- Without placing formal assumptions on the generative model
- Confidence intervals will have correct coverage, even if network information is only crude proxy to true social mechanisms

Kenyan Deworming Experiment[3]

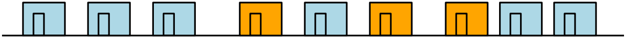


treatment X	1	1	1	0	1	0	0	1	1
# infections Y	2	3	0	22	17	16	22	12	3

- Schools “randomly” selected for de-worming treatment
- Students later measured for parasitic infections
  - Treated: **5.64** infections/school
  - Control: **21.1** infections/school
- Interference: treated students were susceptible to reinfection by untreated ones
  - Bad for T_{spill}, which does well when treated units are well-separated

[3] Miguel and Kremer, *Econometrica* 2004

Kenyan Deworming Experiment[3]



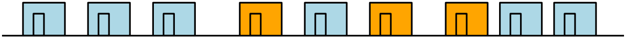
treatment X	1	1	1	0	1	0	0	1	1
# infections Y	2	3	0	22	17	16	22	12	3
counterfactual θ	?	?	?	?	?	?	?	?	?

Notation:

- X_i: treatment of ith school
- Y_i: # of infections at ith school
- θ_i: counterfactual number of infections at school i, if **all** units were treated ("full treatment")

[3] Miguel and Kremer, *Econometrica* 2004

Kenyan Deworming Experiment[3]



treatment X	1	1	1	0	1	0	0	1	1
# infections Y	2	3	0	22	17	16	22	12	3
counterfactual θ	?	?	?	?	?	?	?	?	?

Notation:

- X_i: treatment of ith school
- Y_i: # of infections at ith school
- θ_i: counterfactual number of infections at school i, if **all** units were treated ("full treatment")

Assumption: $\theta_i \leq Y_i$ for all i, i.e., "treatments never hurt"

Goal: Estimate $\bar{\theta} = N^{-1} \sum_i \theta_i$

Note: No other assumptions on interference required

[3] Miguel and Kremer, *Econometrica* 2004

Approach

- Suppose that θ was **observed** for the L treated units:

  A t-test based 95% confidence upper bound for θ:

$$\hat{\theta} + t_{.95}\sqrt{\frac{N-L}{N} \cdot \frac{\hat{\sigma}^2}{L}}, \tag{1}$$

  where $\hat{\theta}$ and $\hat{\sigma}$ are the sample mean and variance:

$$\hat{\theta} = \frac{1}{L}\sum_{\text{treated}} \theta_i \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{L-1}\sum_{\text{treated}} (\theta_i - \hat{\theta})^2$$

- In our setting, θ is **not observed**, but we know $\theta_i \leq Y_i$:

  Upper bound on (1) given by:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{N-L}{N} \cdot \frac{\hat{\sigma}^2}{L}},$$

$$\text{such that } 0 \leq \theta_i \leq Y_i$$ [algorithm for integer θ]

Approach

- Suppose that θ was **observed** for the L treated units:

  A t-test based 95% confidence upper bound for θ:

$$\hat{\theta} + t_{.95}\sqrt{\frac{N-L}{N} \cdot \frac{\hat{\sigma}^2}{L}}, \tag{1}$$

  where $\hat{\theta}$ and $\hat{\sigma}$ are the sample mean and variance:

$$\hat{\theta} = \frac{1}{L}\sum_{\text{treated}} \theta_i \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{L-1}\sum_{\text{treated}} (\theta_i - \hat{\theta})^2$$

- In our setting, θ is **not observed**, but we know $\theta_i \leq Y_i$:

  Upper bound on (1) given by:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{N-L}{N} \cdot \frac{\hat{\sigma}^2}{L}},$$

  such that $0 \leq \theta_i \leq Y_i$ [algorithm for integer θ]

Approach

- Suppose that θ was **observed** for the L treated units:

  A t-test based 95% confidence upper bound for θ:

  $$\hat{\theta} + t_{.95}\sqrt{\frac{N-L}{N} \cdot \frac{\hat{\sigma}^2}{L}}, \tag{1}$$

  where $\hat{\theta}$ and $\hat{\sigma}$ are the sample mean and variance:

  $$\hat{\theta} = \frac{1}{L}\sum_{\text{treated}} \theta_i \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{L-1}\sum_{\text{treated}} (\theta_i - \hat{\theta})^2$$

- In our setting, θ is **not observed**, but we know $\theta_i \leq Y_i$:

  Upper bound on (1) given by:

  $$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{N-L}{N} \cdot \frac{\hat{\sigma}^2}{L}},$$

  such that $0 \leq \theta_i \leq Y_i$ [algorithm for integer θ]

Approach

- Suppose that θ was **observed** for the L treated units:

  A t-test based 95% confidence upper bound for θ:

  $$\hat{\theta} + t_{.95}\sqrt{\frac{N-L}{N} \cdot \frac{\hat{\sigma}^2}{L}}, \tag{1}$$

  where $\hat{\theta}$ and $\hat{\sigma}$ are the sample mean and variance:

  $$\hat{\theta} = \frac{1}{L}\sum_{\text{treated}} \theta_i \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{L-1}\sum_{\text{treated}} (\theta_i - \hat{\theta})^2$$

- In our setting, θ is **not observed**, but we know $\theta_i \leq Y_i$:

  Upper bound on (1) given by:

  $$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{N-L}{N} \cdot \frac{\hat{\sigma}^2}{L}},$$
  $$\text{such that } 0 \leq \theta_i \leq Y_i$$

  [algorithm for integer θ]

Results for Deworming Experiment

With 95% confidence:

- Under full treatment, at most 7.1 infections/school
- Under full control, at least 18.3 infections/school

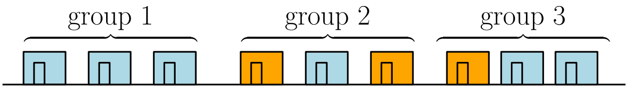
Limitations

Similar to T_{basic}, using $\hat{\theta}$ and $\hat{\sigma}$ does not require any spatial or network information, since:

$$\hat{\theta} = \frac{1}{L} \sum_{\text{treated}} \theta_i \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{L-1} \sum_{\text{treated}} (\theta_i - \hat{\theta})^2$$

- As a result, it can only count direct effects
- No power to detect spillovers
- Cannot rule out the possibility of no interference, so confidence interval must include it
  - [4.2, 7.1] under full treatment for this example

Using Spatial Information

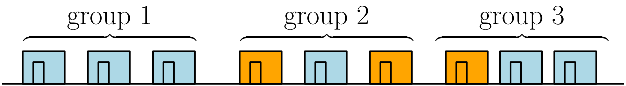


treatment X
\# infections Y
counterfactual θ

- **Pre-treatment**, assign nearby schools into equal-sized groups
- **Declare** that a group is treated if all schools in the group are treated
- **Fact:** Distribution of the treated groups is a **random sample**
- **Use** same upper bound, but with group-level X, Y, and θ:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{n}},$$
$$\text{such that } 0 \leq \theta_i \leq Y_i$$

Using Spatial Information



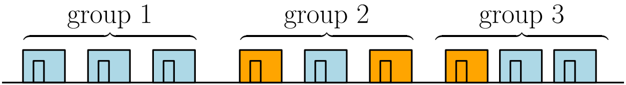
treatment X	1	1	1	0	1	0	0	1	1
# infections Y	2	3	0	22	17	16	22	12	3
counterfactual θ									

- **Pre-treatment**, assign nearby schools into equal-sized groups
- **Declare** that a group is treated if all schools in the group are treated
- **Fact:** Distribution of the treated groups is a **random sample**
- **Use** same upper bound, but with group-level X, Y, and θ:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{n}},$$

$$\text{such that } 0 \le \theta_i \le Y_i$$

Using Spatial Information



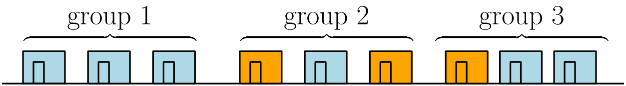
	group 1	group 2	group 3
treatment X	**1**	**0**	**0**
# infections Y	**5**	**55**	**37**
counterfactual θ			

- **Pre-treatment**, assign nearby schools into equal-sized groups
- **Declare** that a group is treated if all schools in the group are treated
- **Fact:** Distribution of the treated groups is a **random sample**
- **Use** same upper bound, but with group-level X, Y, and θ:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{n}},$$

$$\text{such that } 0 \leq \theta_i \leq Y_i$$

Using Spatial Information



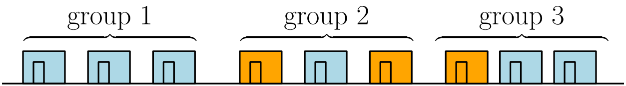
	group 1	group 2	group 3
treatment X	**1**	**0**	**0**
# infections Y	**5**	**55**	**37**
counterfactual θ			

- **Pre-treatment**, assign nearby schools into equal-sized groups
- **Declare** that a group is treated if all schools in the group are treated
- **Fact:** Distribution of the treated groups is a **random sample**
- **Use** same upper bound, but with group-level X, Y, and θ:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{n}},$$

such that $0 \leq \theta_i \leq Y_i$

Using Spatial Information



treatment X	**1**	**0**	**0**
# infections Y	**5**	**55**	**37**
counterfactual θ			

- **Pre-treatment**, assign nearby schools into equal-sized groups
- **Declare** that a group is treated if all schools in the group are treated
- **Fact:** Distribution of the treated groups is a **random sample**
- **Use** same upper bound, but with group-level X, Y, and θ:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{n}},$$
$$\text{such that } 0 \leq \theta_i \leq Y_i$$

Extensions (work in progress)

1. Declare a group to be treated if at least m schools in the group are treated
2. Overlapping groups

For each case, same approach still works:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{n}},$$
$$\text{such that } 0 \leq \theta_i \leq Y_i,$$

but with new formulas for $\hat{\theta}$ and $\hat{\sigma}^2$:

- Two-stage sample
- U-statistic

Tentative result: at most 6.2 infections/school under full treatment

Extensions (work in progress)

1. Declare a group to be treated if at least m schools in the group are treated
2. Overlapping groups

For each case, same approach still works:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{n}},$$
$$\text{such that } 0 \leq \theta_i \leq Y_i,$$

but with new formulas for $\hat{\theta}$ and $\hat{\sigma}^2$:

- Two-stage sample
- U-statistic

Tentative result: at most 6.2 infections/school under full treatment

Extensions (work in progress)

1. Declare a group to be treated if at least m schools in the group are treated
2. Overlapping groups

For each case, same approach still works:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{n}},$$
$$\text{such that } 0 \leq \theta_i \leq Y_i,$$

but with new formulas for $\hat{\theta}$ and $\hat{\sigma}^2$:

- Two-stage sample
- U-statistic

Tentative result: at most 6.2 infections/school under full treatment

Recap

Without spatial or network information, CI includes range given by methods that assume SUTVA:

- Necessary since SUTVA cannot be ruled out
- Any difference in upper bounds could be either because
  - SUTVA might be anti-conservative due to interference, **or**
  - new method might be conservative

With such information, new methods can give improved estimates that rule out hypothesis of no interference

- Without placing formal assumptions on the generative model
- Confidence intervals will have correct coverage, even if network information is only crude proxy to true social mechanisms

Recap

Without spatial or network information, CI includes range given by methods that assume SUTVA:

- Necessary since SUTVA cannot be ruled out
- Any difference in upper bounds could be either because
  - SUTVA might be anti-conservative due to interference, **or**
  - new method might be conservative

With such information, new methods can give improved estimates that rule out hypothesis of no interference

- Without placing formal assumptions on the generative model
- Confidence intervals will have correct coverage, even if network information is only crude proxy to true social mechanisms

Backup Slides

Simple Idea (that doesn't work)

Simple idea: use Y_i directly as an upper bound each θ_i
Problem: error bars may be too small

- Suppose $Y_i = 10, 10, 11, 11, 11$ for untreated units:

$$Y_{\text{avg}} + t_{.95}\sqrt{\frac{\hat{\sigma}_Y^2}{5}} = 11.1$$

- while actually $\theta_i = 0, 10, 11, 11, 11$:

$$\theta_{\text{avg}} + t_{.95}\sqrt{\frac{\hat{\sigma}^2}{5}} = 13.2,$$

- Point estimate using Y is an upper bound, i.e, $Y_{\text{avg}} \geq \theta_{\text{avg}}$
- But confidence interval using Y **decreased**, and loss of coverage results

[back]

What if there are Defiers?

Original Assumption $\theta \leq Y$:

$$\max_{\theta \in \{0,1\}^N} \sum_i \theta_i$$

$$\text{such that } T(X;\theta) \leq t_{.95}(\theta)$$

$$\theta_i \leq Y_i \quad \text{for all } i$$

Under new assumption, treatment effects are

- Nonnegative in aggregate for control
- Arbitrary (including interference) for treated

New formulation gave identical estimate for Facebook experiment

[back]

What if there are Defiers?

Weaker assumption:

$$\max_{\theta \in \{0,1\}^N} \sum_i \theta_i$$

$$\text{such that } T(X;\theta) \leq t_{.95}(\theta)$$

$$\sum_{i:\text{control}} \theta_i \leq \sum_{i:\text{control}} Y_i$$

Under new assumption, treatment effects are

- Nonnegative in aggregate for control
- Arbitrary (including interference) for treated

New formulation gave identical estimate for Facebook experiment

[back]

What if there are Defiers?

Weaker assumption:

$$\max_{\theta \in \{0,1\}^N} \sum_i \theta_i$$

$$\text{such that } T(X;\theta) \le t_{.95}(\theta)$$

$$\sum_{i:\text{control}} \theta_i \le \sum_{i:\text{control}} Y_i$$

Under new assumption, treatment effects are

- Nonnegative in aggregate for control
- Arbitrary (including interference) for treated

New formulation gave identical estimate for Facebook experiment

[back]

Sketch of the Algorithm

- Formulate as an optimization problem:

$$\begin{aligned} & \max_{\theta} \sum_i \theta_i \\ \text{such that} \quad & T(X;\theta) \leq t_\alpha(\theta) \\ & \theta \leq Y. \end{aligned}$$

Sketch of the Algorithm

- Loosen the $T(X;\theta) \le t_\alpha(\theta)$ constraint[4]

$$\begin{aligned} \max_\theta \quad & \sum_i \theta_i \\ \text{such that} \quad & \frac{T(X;\theta) - \mathbb{E}T(X;\theta)}{(\mathrm{Var}\, T(X;\theta))^{1/2}} \le C \\ & \theta \le Y. \end{aligned}$$

[4] This is valid by Chebychev, or even better if T is approximately normal

Sketch of the Algorithm

- Loosen the $T(X;\theta) \le t_\alpha(\theta)$ constraint[4]

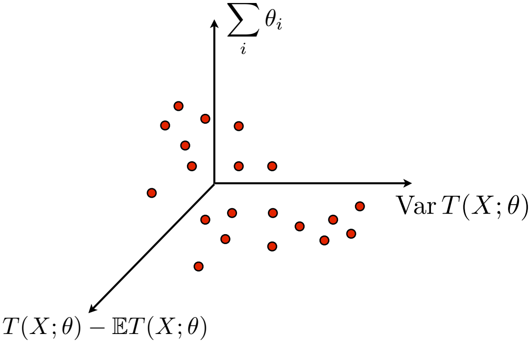
$$\begin{aligned} \max_\theta \quad & \sum_i \theta_i \\ \text{such that} \quad & \frac{T(X;\theta) - \mathbb{E}\,T(X;\theta)}{(\mathrm{Var}\,T(X;\theta))^{1/2}} \le C \\ & \theta \le Y. \end{aligned}$$

- (This is still computationally hard)

[4] This is valid by Chebychev, or even better if T is approximately normal

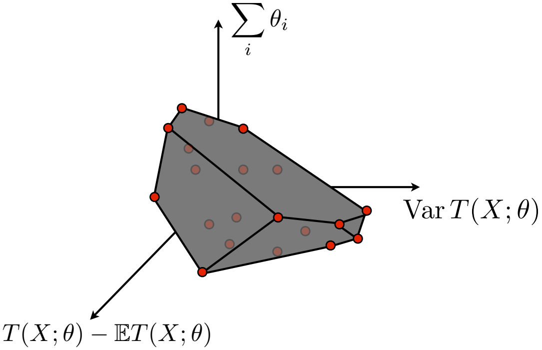
Sketch of the Algorithm

Problem: Number of possible choices for θ satisfying $\theta \leq Y$ is huge



Sketch of the Algorithm

Relaxation: Convex hull is low dimensional and easily searched. Gives upper bound of the objective function



Convex hull can be computed in polynomial time, using method from binary image denoising (which involves Ford-Fulkerson max flow/min cut!) [Grieg, *JRSS B*, 1989] [back]

Solution Method for Integer θ

Want to solve:

$$\max_{\theta} \quad \hat{\theta} + t_{.95}\sqrt{\frac{L-N}{N} \cdot \frac{\hat{\sigma}^2}{L}},$$
$$\text{such that } 0 \leq \theta_i \leq Y_i$$

Exhaustive search: for each possible value of $\hat{\theta}$, find best $\hat{\sigma}$ by solving:

$$\max_{\theta} \quad \sum_{\text{treated}} \theta_i^2$$
$$\text{such that } \frac{1}{L}\sum_{\text{treated}} \theta_i = \hat{\theta}$$
$$0 \leq \theta_i \leq Y_i$$

This is a path planning problem that can be formulated as a dynamic program. [back]