Power Calculations for Randomized Controlled Trials with Auxiliary Observational Data

Jaylin Lowe[1], Charlotte Mann[1], Jiaying Wang[2], Adam Sales[3], and Johann Gagnon-Bartsch[1]

[1]University of Michigan
[2]University of Southern California
[3]Worcester Polytechnic Institute

Educational Data Mining

July 17, 2024

Motivation

- RCTs are the ”gold standard” for determining causality

Gagnon-Bartsch et al., “Precise unbiased estimation in randomized experiments using auxiliary observational data”.

Motivation

- RCTs are the ”gold standard” for determining causality
- Sample sizes in education RCTs are often small, leading to estimates with high variance and low power

Gagnon-Bartsch et al., “Precise unbiased estimation in randomized experiments using auxiliary observational data”.

Motivation

- RCTs are the ”gold standard” for determining causality
- Sample sizes in education RCTs are often small, leading to estimates with high variance and low power
- We can improve precision by incorporating auxiliary data[1]

Gagnon-Bartsch et al., “Precise unbiased estimation in randomized experiments using auxiliary observational data”.

Motivation

- RCTs are the ”gold standard” for determining causality
- Sample sizes in education RCTs are often small, leading to estimates with high variance and low power
- We can improve precision by incorporating auxiliary data[1]
- With this method, sample sizes can be smaller with the same power

Gagnon-Bartsch et al., “Precise unbiased estimation in randomized experiments using auxiliary observational data”.

Motivation

- RCTs are the ”gold standard” for determining causality
- Sample sizes in education RCTs are often small, leading to estimates with high variance and low power
- We can improve precision by incorporating auxiliary data[1]
- With this method, sample sizes can be smaller with the same power

Main Question:

If you are designing an experiment and intend to use auxiliary data to improve precision in the estimates, how do you determine what sample size you will need?

Gagnon-Bartsch et al., “Precise unbiased estimation in randomized experiments using auxiliary observational data”.

Requirements

This method may be useful to you if you are designing an RCT and you:

1. Need an idea of what sample size you should use

Requirements

This method may be useful to you if you are designing an RCT and you:

1. Need an idea of what sample size you should use
2. Are considering using auxiliary data and have access to an auxiliary dataset that:

Requirements

This method may be useful to you if you are designing an RCT and you:

1. Need an idea of what sample size you should use
2. Are considering using auxiliary data and have access to an auxiliary dataset that:
    - Has the same outcome of interest as your RCT

Requirements

This method may be useful to you if you are designing an RCT and you:

1. Need an idea of what sample size you should use
2. Are considering using auxiliary data and have access to an auxiliary dataset that:
    - Has the same outcome of interest as your RCT
    - Has covariates

Requirements

This method may be useful to you if you are designing an RCT and you:

1. Need an idea of what sample size you should use
2. Are considering using auxiliary data and have access to an auxiliary dataset that:
    - Has the same outcome of interest as your RCT
    - Has covariates
    - Has a large number of observations

Requirements

This method may be useful to you if you are designing an RCT and you:

1. Need an idea of what sample size you should use
2. Are considering using auxiliary data and have access to an auxiliary dataset that:
    - Has the same outcome of interest as your RCT
    - Has covariates
    - Has a large number of observations
    - Somewhat resembles your RCT population

Problem

- Decrease in necessary sample size using this method is determined by how predictive a model fit on the auxiliary data is for the RCT sample

Problem

- Decrease in necessary sample size using this method is determined by how predictive a model fit on the auxiliary data is for the RCT sample
- Model is highly predictive $\rightarrow$ RCT estimates are more precise $\rightarrow$ RCT can be smaller with the same power

Problem

- Decrease in necessary sample size using this method is determined by how predictive a model fit on the auxiliary data is for the RCT sample
- Model is highly predictive $\rightarrow$ RCT estimates are more precise $\rightarrow$ RCT can be smaller with the same power
- Model is not predictive at all $\rightarrow$ No gain in precision $\rightarrow$ RCT sample size cannot be decreased

Problem

- Decrease in necessary sample size using this method is determined by how predictive a model fit on the auxiliary data is for the RCT sample
- Model is highly predictive $\rightarrow$ RCT estimates are more precise $\rightarrow$ RCT can be smaller with the same power
- Model is not predictive at all $\rightarrow$ No gain in precision $\rightarrow$ RCT sample size cannot be decreased
- We don't have the RCT sample ahead of time!

Idea

- Break the auxiliary data up into subgroups

Idea

- Break the auxiliary data up into subgroups
- For each subgroup:

Idea

- Break the auxiliary data up into subgroups
- For each subgroup:
  - Treat the subgroup as the RCT and the rest of the data as the auxiliary data

Idea

- Break the auxiliary data up into subgroups
- For each subgroup:
  - Treat the subgroup as the RCT and the rest of the data as the auxiliary data
  - Calculate required sample size under this framework

Idea

- Break the auxiliary data up into subgroups
- For each subgroup:
  - Treat the subgroup as the RCT and the rest of the data as the auxiliary data
  - Calculate required sample size under this framework
- Use entire range of sample sizes generated from different subgroups to get an idea of necessary sample size.

Example Application: CTA Study

- New algebra curriculum with personalized tutoring software

Example Application: CTA Study

- New algebra curriculum with personalized tutoring software
- Schools randomized to implement this new curriculum or stick with the standard

Example Application: CTA Study

- New algebra curriculum with personalized tutoring software
- Schools randomized to implement this new curriculum or stick with the standard
- Outcome variable: passing rate on subsequent standardized tests

Example Application: CTA Study

- New algebra curriculum with personalized tutoring software
- Schools randomized to implement this new curriculum or stick with the standard
- Outcome variable: passing rate on subsequent standardized tests
- We have access to school level standardized test scores and other school level information for schools in Texas

Example Application: CTA Study

- New algebra curriculum with personalized tutoring software
- Schools randomized to implement this new curriculum or stick with the standard
- Outcome variable: passing rate on subsequent standardized tests
- We have access to school level standardized test scores and other school level information for schools in Texas
- Create subgroups based on best case/worst case scenario

Example Results

	Auxiliary Data?	
Decile	**Yes**	**No**
1	19	92
2	47	200
3	50	272
4	81	424
5	108	382
6	113	516
7	123	497
8	191	622
9	370	1094
10	998	1829

Shiny App

- Interactive interface for exploring this method and obtaining a range of reasonable sample sizes

Shiny App

- Interactive interface for exploring this method and obtaining a range of reasonable sample sizes
- Users can:

Shiny App

- Interactive interface for exploring this method and obtaining a range of reasonable sample sizes
- Users can:
  - Specify how subgroups should be determined, especially if there are particular covariates of interest

Shiny App

- Interactive interface for exploring this method and obtaining a range of reasonable sample sizes
- Users can:
  - Specify how subgroups should be determined, especially if there are particular covariates of interest
  - Explore covariate values and patterns in different subgroups

Shiny App

- Interactive interface for exploring this method and obtaining a range of reasonable sample sizes
- Users can:
  - Specify how subgroups should be determined, especially if there are particular covariates of interest
  - Explore covariate values and patterns in different subgroups
  - Specify other parameters in the sample size calculation

Shiny App

- Interactive interface for exploring this method and obtaining a range of reasonable sample sizes
- Users can:
  - Specify how subgroups should be determined, especially if there are particular covariates of interest
  - Explore covariate values and patterns in different subgroups
  - Specify other parameters in the sample size calculation
  - Compare sample sizes with or without auxiliary data

Shiny App

- Interactive interface for exploring this method and obtaining a range of reasonable sample sizes
- Users can:
  - Specify how subgroups should be determined, especially if there are particular covariates of interest
  - Explore covariate values and patterns in different subgroups
  - Specify other parameters in the sample size calculation
  - Compare sample sizes with or without auxiliary data
- Shiny app can be downloaded from `https://github.com/jaylinlowe/dRCTpower`

Questions?

Contact: jaylincl@umich.edu