A Theory of Experimenters

Abhijit Banerjee (MIT)

with Sylvain Chassang (NYU),
and Erik Snowberg (UBC),
and Sergio Montero (University of Rochester),

PSE
November 2017

Motivation

- Proliferation of experiments (academia, business, public policy) has been accompanied by debates about best practices

Motivation

- Proliferation of experiments (academia, business, public policy) has been accompanied by debates about best practices
  - Blocking, clustering, stratification

Motivation

- Proliferation of experiments (academia, business, public policy) has been accompanied by debates about best practices
  - Blocking, clustering, stratification
  - Pros and cons of randomization

Motivation

- Proliferation of experiments (academia, business, public policy) has been accompanied by debates about best practices
  - Blocking, clustering, stratification
  - Pros and cons of randomization
  - Pre-registration of designs and pre-analysis plans

Motivation

- Proliferation of experiments (academia, business, public policy) has been accompanied by debates about best practices
  - Blocking, clustering, stratification
  - Pros and cons of randomization
  - Pre-registration of designs and pre-analysis plans
- Theory should help resolve these debates

Motivation

- Proliferation of experiments (academia, business, public policy) has been accompanied by debates about best practices
  - Blocking, clustering, stratification
  - Pros and cons of randomization
  - Pre-registration of designs and pre-analysis plans
- Theory should help resolve these debates
- However, standard models of information acquisition fail to explain key feature of experimental practice: randomization

Motivation

- Proliferation of experiments (academia, business, public policy) has been accompanied by debates about best practices
  - Blocking, clustering, stratification
  - Pros and cons of randomization
  - Pre-registration of designs and pre-analysis plans
- Theory should help resolve these debates
- However, standard models of information acquisition fail to explain key feature of experimental practice: randomization
  - RCTs are mixed strategies over experimental assignments $\rightarrow$ never strictly optimal for Bayesian decision maker

Goal

- Propose a decision-theoretic framework for experiment design that:

Goal

- Propose a decision-theoretic framework for experiment design that:
    - correctly captures the preferences revealed by real-life experimentation

Goal

- Propose a decision-theoretic framework for experiment design that:
  - correctly captures the preferences revealed by real-life experimentation
  - provides insight into open problems for experimental practice

Outline of the Talk

- *Building plausible model of experimentation*

  Fact: people really care about randomization

Outline of the Talk

- *Building plausible model of experimentation*

  Fact: people really care about randomization

  Bayesians don't randomize (Kasy 2013)
  Ambiguity averse experimenters might

Outline of the Talk

- *Building plausible model of experimentation*

  Fact: people really care about randomization

  Bayesians don't randomize (Kasy 2013)
  Ambiguity averse experimenters might

- *Use it to address re-randomization*

  Fact: people often don't stick with their first randomization (Bruhn and McKenzie 2009)

Outline of the Talk

- *Building plausible model of experimentation*

  Fact: people really care about randomization

  Bayesians don't randomize (Kasy 2013)
  Ambiguity averse experimenters might

- *Use it to address re-randomization*

  Fact: people often don't stick with their first randomization (Bruhn and McKenzie 2009)

  It does reduce robustness, but very slowly

An example: a voucher experiment

- A school district superintendent wants to do an experiment
- Her prior puts a lot of weight on the idea that private schools are all about selection and that private school students will do equally well in private and public schools
- However she allows that there is some probability that private schools are better and that all children would do much better there
- She has one slot in a private school: how should she allocate it?
- Clearly giving it to a poor child maximizes her learning.

The experiment continues

- Now suppose the superintendent assigns one more child to the experiment.
- The best design under her priors will be to assign a rich child to the public school and a poor child to a private school.
- No randomization
- Not balanced. A Bayesian may not want balance.
    - Contrast with Kasy (2014)
- Even if she only had two children who were both poor for the experiment, she has no reason to randomize.

Setup

- Population indexed by i
  - Covariates $x_i \in X$ observed
  - Distributed according to $q(x)$

Setup

- Population indexed by i
  - Covariates $x_i \in X$ observed
  - Distributed according to $q(x)$
- Treatment $\tau \in \{0, 1\}$, potential outcome $Y_i^\tau \in \{0, 1\}$

Setup

- Population indexed by i
  - Covariates $x_i \in X$ observed
  - Distributed according to $q(x)$
- Treatment $\tau \in \{0, 1\}$, potential outcome $Y_i^\tau \in \{0, 1\}$
- $p_x^\tau \equiv \mathrm{Prob}(Y_i^\tau = 1 | x_i = x)$,

Setup

- Population indexed by i
  - Covariates $x_i \in X$ observed
  - Distributed according to $q(x)$
- Treatment $\tau \in \{0, 1\}$, potential outcome $Y_i^\tau \in \{0, 1\}$
- $p_x^\tau \equiv \text{Prob}(Y_i^\tau = 1 | x_i = x)$,
  state of the world $p = (p_x^\tau)_{x,\tau} \in \mathcal{P}$

Setup

- Population indexed by i
  - Covariates $x_i \in X$ observed
  - Distributed according to $q(x)$
- Treatment $\tau \in \{0,1\}$, potential outcome $Y_i^\tau \in \{0,1\}$
- $p_x^\tau \equiv \mathrm{Prob}(Y_i^\tau = 1 | x_i = x)$,
  state of the world $p = (p_x^\tau)_{x,\tau} \in \mathcal{P}$
- Policy decision $a \in \{0,1\}$: treat everybody or not

Setup

- Population indexed by i
  - Covariates $x_i \in X$ observed
  - Distributed according to $q(x)$
- Treatment $\tau \in \{0,1\}$, potential outcome $Y_i^\tau \in \{0,1\}$
- $p_x^\tau \equiv \mathrm{Prob}(Y_i^\tau = 1 | x_i = x)$,
  state of the world $p = (p_x^\tau)_{x,\tau} \in \mathcal{P}$
- Policy decision $a \in \{0,1\}$: treat everybody or not
- Decision maker's payoff

$$u(a,p) \equiv \sum_{x \in X} q(x) p_x^a.$$

Setup

- Population indexed by i
  - Covariates $x_i \in X$ observed
  - Distributed according to $q(x)$
- Treatment $\tau \in \{0, 1\}$, potential outcome $Y_i^\tau \in \{0, 1\}$
- $p_x^\tau \equiv \mathrm{Prob}(Y_i^\tau = 1 | x_i = x)$,
  state of the world $p = (p_x^\tau)_{x,\tau} \in \mathcal{P}$
- Policy decision $a \in \{0, 1\}$: treat everybody or not
- Decision maker's payoff

$$u(a, p) \equiv \sum_{x \in X} q(x) p_x^a.$$

- E.g., vaccinate school children or not, reorganize production lines

Experiments

- Decision maker informs policy choice by running an experiment on $N(< 2|X|)$ subjects

Experiments

- Decision maker informs policy choice by running an experiment on $N(< 2|X|)$ subjects
- Experiment design is mixed strategy $\mathcal{E} \in \Delta(E)$ over experimental sample and treatment tuples $e = (x_i, \tau_i)_{i \in \{1,\ldots,N\}} \in E$

Experiments

- Decision maker informs policy choice by running an experiment on $N(< 2|X|)$ subjects
- Experiment design is mixed strategy $\mathcal{E} \in \Delta(E)$ over experimental sample and treatment tuples $e = (x_i, \tau_i)_{i \in \{1,\ldots,N\}} \in E$
- Generates outcome data $y = (y_i)_{i \in \{1,\ldots,N\}} \in \mathcal{Y}$

Experiments

- Decision maker informs policy choice by running an experiment on $N(< 2|X|)$ subjects
- Experiment design is mixed strategy $\mathcal{E} \in \Delta(E)$ over experimental sample and treatment tuples $e = (x_i, \tau_i)_{i \in \{1,\dots,N\}} \in E$
- Generates outcome data $y = (y_i)_{i \in \{1,\dots,N\}} \in \mathcal{Y}$
- Allocation rule $\alpha : E \times \mathcal{Y} \to \Delta(\{0,1\})$

Natural Model

Subjective expected utility maximizer (Bayesian)

- Picks $\mathcal{E}, \alpha$ solving

  $$\max_{\mathcal{E},\alpha} \mathbb{E}_h[u(\alpha, p)]$$

  for prior $h \in \Delta(\mathcal{P})$ over state of the world p

Not a Great Positive Model

Proposition 1.

Bayesian $\Rightarrow$ *optimal experiment* $\mathcal{E}^*$ *deterministic.*

Not a Great Positive Model

Proposition 1.

Bayesian $\Rightarrow$ optimal experiment $\mathcal{E}^*$ deterministic.

For generic priors, deterministic experiment strictly optimal.

Not a Great Positive Model

Proposition 1.

Bayesian $\Rightarrow$ optimal experiment $\mathcal{E}^*$ deterministic.

For generic priors, deterministic experiment strictly optimal.

Why?

- Randomization is a mixed strategy
- Pure strategies weakly optimal for expected utility maximizer

Not a Great Positive Model

Proposition 1.

Bayesian $\Rightarrow$ optimal experiment $\mathcal{E}^*$ deterministic.

For generic priors, deterministic experiment strictly optimal.

Why?

- Randomization is a mixed strategy
- Pure strategies weakly optimal for expected utility maximizer

  Payoff from experiment $\mathcal{E}$:

$$\mathbb{E}_{e,y\sim\mathcal{E}} \max_{a\in\{0,1\}} \mathbb{E}_h[u(p,a)|e,y]$$

Not Completely Off Either

- Imagine you are a presidential candidate, a few days before a key media appearance

Not Completely Off Either

- Imagine you are a presidential candidate, a few days before a key media appearance
- Two speeches: first pretty classic, second tries out new points

Not Completely Off Either

- Imagine you are a presidential candidate, a few days before a key media appearance
- Two speeches: first pretty classic, second tries out new points
- Opportunity to experiment by giving one speech at a campaign meeting a few days before

Not Completely Off Either

- Imagine you are a presidential candidate, a few days before a key media appearance
- Two speeches: first pretty classic, second tries out new points
- Opportunity to experiment by giving one speech at a campaign meeting a few days before
- Do you pick the location randomly so that it's representative? Do you randomize which speech you give?

Not Completely Off Either

- Imagine you are a presidential candidate, a few days before a key media appearance
- Two speeches: first pretty classic, second tries out new points
- Opportunity to experiment by giving one speech at a campaign meeting a few days before
- Do you pick the location randomly so that it's representative? Do you randomize which speech you give?
- Obviously, no; you'll pick most informative experiment: *prior gives you a comparison point even with a single outcome*

Not Completely Off Either

- Imagine you are a presidential candidate, a few days before a key media appearance
- Two speeches: first pretty classic, second tries out new points
- Opportunity to experiment by giving one speech at a campaign meeting a few days before
- Do you pick the location randomly so that it's representative? Do you randomize which speech you give?
- Obviously, no; you'll pick most informative experiment: *prior gives you a comparison point even with a single outcome*
- With a prior, even with two meetings, you might give the same speech at both

Ambiguity Averse Experimentation

- Decision maker picks $\mathcal{E}, \alpha$ solving

$$\max_{\mathcal{E},\alpha} \ \lambda \mathbb{E}_{h_0}[u(\alpha, p)] + (1-\lambda) \min_{h \in H} \mathbb{E}_h[u(\alpha, p)]$$

Ambiguity Averse Experimentation

- Decision maker picks $\mathcal{E}, \alpha$ solving

$$\max_{\mathcal{E},\alpha} \ \lambda \mathbb{E}_{h_0}[u(\alpha,p)] + (1-\lambda)\min_{h\in H}\mathbb{E}_h[u(\alpha,p)]$$

- Interpretation: subjective prior h_0, constrained by adversarial audience

Ambiguity Averse Experimentation

- Decision maker picks $\mathcal{E}, \alpha$ solving

$$\max_{\mathcal{E},\alpha} \; \lambda \mathbb{E}_{h_0}[u(\alpha, p)] + (1-\lambda) \min_{h \in H} \mathbb{E}_h[u(\alpha, p)]$$

- Interpretation: subjective prior h_0, constrained by adversarial audience

Assumption 1 (Limited Extrapolation).

For all realized experiments $\mathbf{e}$*, there exists an adversarial prior* $\mathbf{h}$ *such that optimal decisions conditional on data are bounded away from first best*
(i.e., even with infinite data, there is room for learning)

Ambiguity Averse Experimentation

- Decision maker picks $\mathcal{E}, \alpha$ solving

$$\max_{\mathcal{E},\alpha} \; \lambda \mathbb{E}_{h_0}[u(\alpha, p)] + (1-\lambda) \min_{h \in H} \mathbb{E}_h[u(\alpha, p)]$$

- Interpretation: subjective prior h_0, constrained by adversarial audience

Assumption 1 (Limited Extrapolation).

For all realized experiments $\mathbf{e}$, *there exists an adversarial prior* $\mathbf{h}$ *such that optimal decisions conditional on data are bounded away from first best*
(i.e., even with infinite data, there is room for learning)

Can be dispensed with if DM exhibits regret aversion

Optimal Maxmin Design

Proposition 2.

For fixed N *and generic* h_0,

Optimal Maxmin Design

Proposition 2.

For fixed N and generic h_0,

if $H \to \{h_0\}$ (audience not adversarial)
or $\lambda \to 1$ (don't care about convincing others),

Optimal Maxmin Design

Proposition 2.

For fixed N and generic h_0,

if $H \to \{h_0\}$ (audience not adversarial)
or $\lambda \to 1$ (don't care about convincing others),

then optimal experiment deterministic and Bayesian optimal for h_0

Optimal Maxmin Design

Proposition 3.

Take $\lambda \in (0, 1)$ *as given.*

Optimal Maxmin Design

Proposition 3.

Take $\lambda \in (0,1)$ as given.

(i) Optimal experiment (e.g., std RCT) guarantees

$$\max_{\alpha} \min_{h\in H} \mathbb{E}_{h,\mathcal{E}} \left[u(p, \alpha(e,y)) \right] > \min_{h\in H} \mathbb{E}_h \left(\max_{a\in\{0,1\}} u(p,a) \right) - \sqrt{\frac{\ln 2}{N}}.$$

Optimal Maxmin Design

Proposition 3.

Take $\lambda \in (0,1)$ as given.

(i) Optimal experiment (e.g., std RCT) guarantees

$$\max_{\alpha} \min_{h \in H} \mathbb{E}_{h,\mathcal{E}} \left[u(p, \alpha(e,y)) \right] > \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \sqrt{\frac{\ln 2}{N}}.$$

(ii) Deterministic experiments are bounded away from (maxmin) efficiency:

$$\max_{\alpha} \min_{h \in H} \mathbb{E}_{h,e} \left[u(p, \alpha(e,y)) \right] < \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \xi.$$

Optimal Maxmin Design

Proposition 3.

Take $\lambda \in (0,1)$ as given.

(i) Optimal experiment (e.g., std RCT) guarantees

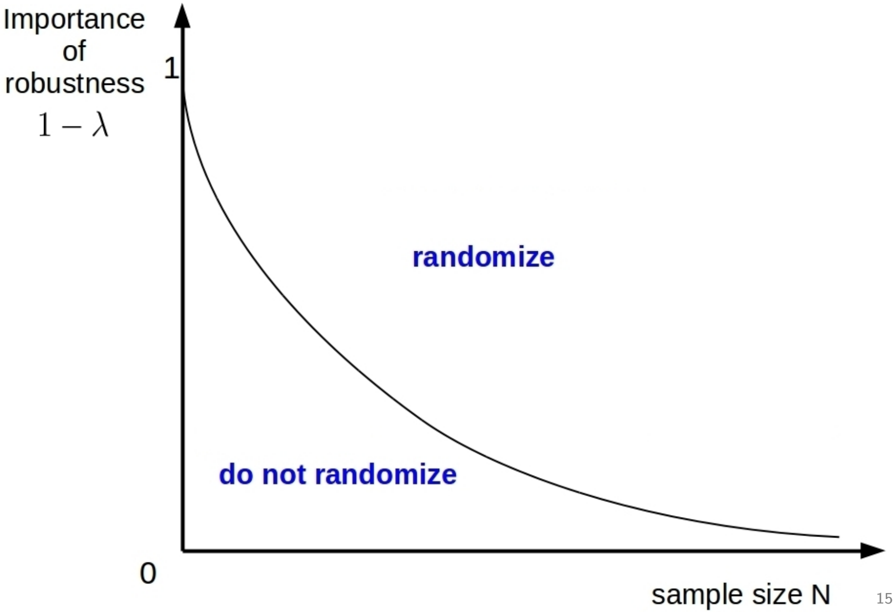
$$\max_{\alpha} \min_{h \in H} \mathbb{E}_{h,\mathcal{E}} \left[u(p, \alpha(e,y)) \right] > \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \sqrt{\frac{\ln 2}{N}}.$$

(ii) Deterministic experiments are bounded away from (maxmin) efficiency:

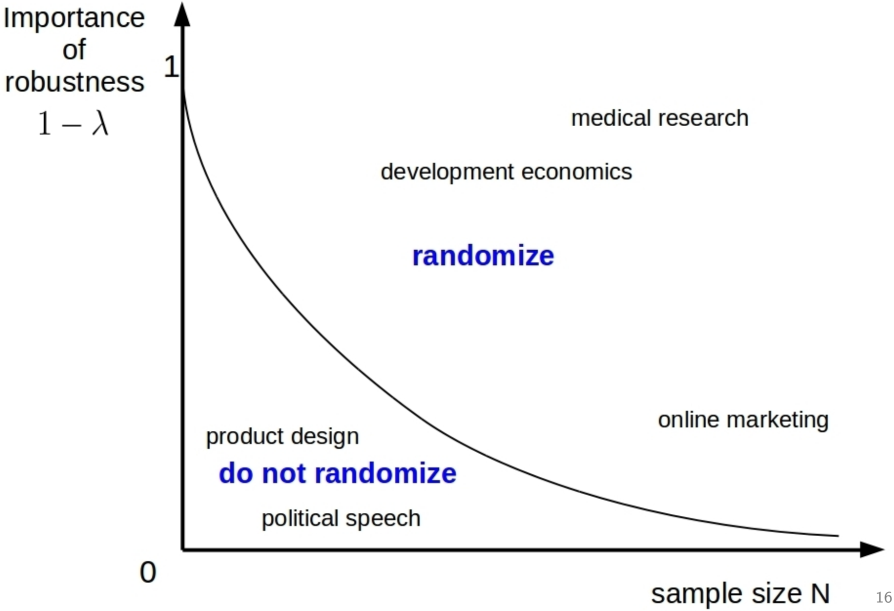
$$\max_{\alpha} \min_{h \in H} \mathbb{E}_{h,e} \left[u(p, \alpha(e,y)) \right] < \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \xi.$$

As sample size N gets large, **optimal experiment is random**

Design Choice



Design Choice



Intuition, Other Implications

- Playing a zero-sum game against nature

Intuition, Other Implications

- Playing a zero-sum game against nature
- Anything that makes you (as the experimenter) predictable will be matched with an unfavorable prior ($h \in H$) that exploits that predictability

Intuition, Other Implications

- Playing a zero-sum game against nature
- Anything that makes you (as the experimenter) predictable will be matched with an unfavorable prior ($h \in H$) that exploits that predictability
- Implication: quasi-randomization not sufficient
  - Randomizing by time of day; see Green and Tusicisny, 2012, for a critique
  - Miguel and Kremer, 2004; see Deaton, 2010 for a critique

Intuition, Other Implications

- Playing a zero-sum game against nature
- Anything that makes you (as the experimenter) predictable will be matched with an unfavorable prior ($h \in H$) that exploits that predictability
- Implication: quasi-randomization not sufficient
  - Randomizing by time of day; see Green and Tusicisny, 2012, for a critique
  - Miguel and Kremer, 2004; see Deaton, 2010 for a critique
- Implication: RCTs offer near optimal alternative to complexity of solving decision maker's problem exactly, which requires reliably eliciting beliefs (priors)

Balance

- Experimenters often want a "balanced" assignment

Balance

- Experimenters often want a “balanced” assignment
- Stratification or blocking often used
  - Hard to do with multiple or continuous variables
  - Related to “curse of dimensionality”
  - In the U.S., Gender x Race x Age (x groups) x Education (y groups) $= 10xy$ bins

Balance

- Experimenters often want a “balanced” assignment
- Stratification or blocking often used
  - Hard to do with multiple or continuous variables
  - Related to “curse of dimensionality”
  - In the U.S., Gender x Race x Age (x groups) x Education (y groups) $= 10xy$ bins
- However, these algorithms create predictable assignments
  - Is this a weakness?
  - Surprisingly, yes: “You picked the wrong variables to block on”

What If You Get an Unbalanced Sample?

- Max-min preferences not dynamically consistent:
  What if assignment very unbalanced? Re-randomize?

What If You Get an Unbalanced Sample?

- Max-min preferences not dynamically consistent:
  What if assignment very unbalanced? Re-randomize?

- Not a problem for Bayesian learner:

$$\mathrm{Prob}(p|e, y, e \text{ picked randomly}) = \mathrm{Prob}(p|e, y, e \text{ selected})$$

$\rightarrow$ process of experiment design irrelevant

What If You Get an Unbalanced Sample?

- Max-min preferences not dynamically consistent:
  What if assignment very unbalanced? Re-randomize?

- Not a problem for Bayesian learner:

  $$\mathrm{Prob}(p|e, y, e \text{ picked randomly}) = \mathrm{Prob}(p|e, y, e \text{ selected})$$

  $\rightarrow$ process of experiment design irrelevant

- Is this a problem for robustness? Can we quantify it?

Re-Randomization

Model can be written as

$$\max_{\mathcal{E},\alpha} \; \lambda \underbrace{\mathbb{E}_{\mathcal{E}}[B(e,\alpha)]}_{\text{subjective balance}} + (1-\lambda) \underbrace{R(\mathcal{E},\alpha)}_{\text{robustness}}$$

Re-Randomization

Model can be written as

$$\max_{\mathcal{E},\alpha} \lambda \underbrace{\mathbb{E}_{\mathcal{E}}[B(e,\alpha)]}_{\text{subjective balance}} + (1-\lambda) \underbrace{R(\mathcal{E},\alpha)}_{\text{robustness}}$$

K **re-randomization**:

1. Fixed sample of xs drawn according to pop. dist. $q \in \Delta(X)$, independently draw K assignments $\{e_1, \cdots, e_K\}$ (prob. treatment $=$ .5)

Re-Randomization

Model can be written as

$$\max_{\mathcal{E},\alpha} \lambda \underbrace{\mathbb{E}_{\mathcal{E}}[B(e,\alpha)]}_{\text{subjective balance}} + (1-\lambda) \underbrace{R(\mathcal{E},\alpha)}_{\text{robustness}}$$

K **re-randomization**:

1. Fixed sample of xs drawn according to pop. dist. $q \in \Delta(X)$, independently draw K assignments $\{e_1, \cdots, e_K\}$ (prob. treatment $= .5$)
2. Select assignment $e_K^* \in \arg\max_{e \in \{e_1, \cdots, e_K\}} \max_\alpha B(e,\alpha)$ maximizing Bayesian expected utility

Re-Randomization

Model can be written as

$$\max_{\mathcal{E},\alpha} \; \lambda \underbrace{\mathbb{E}_{\mathcal{E}}[B(e,\alpha)]}_{\text{subjective balance}} + (1-\lambda) \underbrace{R(\mathcal{E},\alpha)}_{\text{robustness}}$$

K **re-randomization**:

1. Fixed sample of xs drawn according to pop. dist. $q \in \Delta(X)$, independently draw K assignments $\{e_1, \cdots, e_K\}$ (prob. treatment = .5)
2. Select assignment $e_K^* \in \arg\max_{e \in \{e_1, \cdots, e_K\}} \max_\alpha B(e,\alpha)$ maximizing Bayesian expected utility
3. Run experiment e_K^*

Re-Randomization

Model can be written as

$$\max_{\mathcal{E},\alpha} \; \lambda \underbrace{\mathbb{E}_{\mathcal{E}}[B(e,\alpha)]}_{\text{subjective balance}} + (1-\lambda) \underbrace{R(\mathcal{E},\alpha)}_{\text{robustness}}$$

K **re-randomization**:

1. Fixed sample of xs drawn according to pop. dist. $q \in \Delta(X)$, independently draw K assignments $\{e_1, \cdots, e_K\}$ (prob. treatment $=$ .5)
2. Select assignment $e_K^* \in \arg\max_{e \in \{e_1, \cdots, e_K\}} \max_\alpha B(e,\alpha)$ maximizing Bayesian expected utility
3. Run experiment e_K^*
4. Choose policy according to $\alpha^* = \arg\max_{a \in \{0,1\}} \overline{y}^a - \overline{y}^{1-a}$

The Tradeoff of Re-Randomization

Improves balance

- $B(e_K^*, \alpha^*)$ monotonically increasing in K

The Tradeoff of Re-Randomization

Improves balance

- $B(e_K^*, \alpha^*)$ monotonically increasing in K

Proposition 4 (negative impact on robustness).
There exists $\rho > 0$ such that, for all N, if $K \geq 2^N$, then

$$\max_{\alpha} \min_{h \in H} \mathbb{E}_{h, \mathcal{E}_K} \left[u(p, \alpha(e, y)) \right] < \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p, a) \right) - \rho.$$

How Large Are the Costs?

Proposition 5 (cost of rerandomization small).

A K-rerandomized experiment $\mathcal{E}_K$ guarantees

$$\min_{h \in H} \mathbb{E}_{h,\mathcal{E}_K}\left[u(p, \alpha(e, y))\right] \geq \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p, a) \right) - \sqrt{\frac{\ln(K)}{N/2}}$$

How Large Are the Costs?

Proposition 5 (cost of rerandomization small).

A K-rerandomized experiment $\mathcal{E}_K$ guarantees

$$\min_{h \in H} \mathbb{E}_{h,\mathcal{E}_K}\left[u(p, \alpha(e,y))\right] \geq \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \sqrt{\frac{\ln(K)}{N/2}}$$

Proposal 1.

Set $K \leq N$, for instance $K = \min\{N, 100\}$

How Large Are the Costs?

Proposition 5 (cost of rerandomization small).

A K-rerandomized experiment $\mathcal{E}_K$ guarantees

$$\min_{h \in H} \mathbb{E}_{h,\mathcal{E}_K}\left[u(p, \alpha(e,y))\right] \geq \min_{h \in H} \mathbb{E}_h \left(\max_{a \in \{0,1\}} u(p,a) \right) - \sqrt{\frac{\ln(K)}{N/2}}$$

Proposal 1.

Set $K \leq N$, for instance $K = \min\{N, 100\}$

Remark 1.

Bound remains valid regardless of objective function $B(e)$, can even choose objective ex post

Numerical Assessment

- Consider likelihood with which K-rerandomized trial generates assignment in top z^{th} quantile of balance

Numerical Assessment

- Consider likelihood with which K-rerandomized trial generates assignment in top z^{th} quantile of balance
- That likelihood is equal to $1 - (1 - z)^K$

Numerical Assessment

- Consider likelihood with which K-rerandomized trial generates assignment in top z^{th} quantile of balance
- That likelihood is equal to $1 - (1 - z)^K$

K	10	50	100	250	500	1000
$\sqrt{\log(K)}$	1.52	1.97	2.15	2.35	2.49	2.63
odds top 5% bal.	0.4	0.92	0.99	1.0	1.0	1.0
odds top 1% bal.	0.1	0.39	0.63	0.92	0.99	1.0

What About Standard Errors?

- Hypothesis testing (using t- or z-stats) is not Bayesian, not due to risk aversion

What About Standard Errors?

- Hypothesis testing (using t- or z-stats) is not Bayesian, not due to risk aversion
- Tetenov (2012) shows that reference dependence (loss aversion or status quo bias) can rationalize hypothesis testing

What About Standard Errors?

- Hypothesis testing (using t- or z-stats) is not Bayesian, not due to risk aversion
- Tetenov (2012) shows that reference dependence (loss aversion or status quo bias) can rationalize hypothesis testing

Proposition 6.

All of our results extend to using reference dependent preferences

What About Standard Errors?

- Hypothesis testing (using t- or z-stats) is not Bayesian, not due to risk aversion
- Tetenov (2012) shows that reference dependence (loss aversion or status quo bias) can rationalize hypothesis testing

Proposition 6.
All of our results extend to using reference dependent preferences

- Morgan and Rubin (2012) show that re-randomization increases precision of estimated treatment effect in linear Gaussian model

What About Standard Errors?

- Hypothesis testing (using t- or z-stats) is not Bayesian, not due to risk aversion
- Tetenov (2012) shows that reference dependence (loss aversion or status quo bias) can rationalize hypothesis testing

Proposition 6.

All of our results extend to using reference dependent preferences

- Morgan and Rubin (2012) show that re-randomization increases precision of estimated treatment effect in linear Gaussian model
- Bungi, Canay, and Shaikh (2016) show this more generally for balanced assignment rules (i.e., symmetric stratification)

Alternative Procedure

- Could consider defining a set of acceptable assignments and randomize over them:
  e.g., require sufficiently good balance under some distance (Morgan and Rubin 2012)

Alternative Procedure

- Could consider defining a set of acceptable assignments and randomize over them:
  e.g., require sufficiently good balance under some distance (Morgan and Rubin 2012)

- Difficulty: is the balance requirement hard to satisfy?
  If so, then near deterministic assignment

Alternative Procedure

- Could consider defining a set of acceptable assignments and randomize over them:
  e.g., require sufficiently good balance under some distance (Morgan and Rubin 2012)

- Difficulty: is the balance requirement hard to satisfy?
  If so, then near deterministic assignment

- Can be addressed in our framework: set $B(e) \equiv \mathbf{1}_{\text{balance}>\underline{b}}$

Alternative Procedure

- Could consider defining a set of acceptable assignments and randomize over them:
  e.g., require sufficiently good balance under some distance (Morgan and Rubin 2012)
- Difficulty: is the balance requirement hard to satisfy?
  If so, then near deterministic assignment
- Can be addressed in our framework: set $B(e) \equiv \mathbf{1}_{\text{balance}>\underline{b}}$
- If probability that a random assignment is balanced is very small, then procedure above is akin to setting K very high

Takeaway

- Ambiguity-averse experimentation as a plausible model for a range of behavior

  With small samples, we need to be Bayesian — learning necessarily subjective

Takeaway

- Ambiguity-averse experimentation as a plausible model for a range of behavior

  With small samples, we need to be Bayesian — learning necessarily subjective

Takeaway

- Ambiguity-averse experimentation as a plausible model for a range of behavior

  With small samples, we need to be Bayesian — learning necessarily subjective

- Re-randomization does involve a tradeoff, but cost is small

Takeaway

- Ambiguity-averse experimentation as a plausible model for a range of behavior

  With small samples, we need to be Bayesian — learning necessarily subjective

- Re-randomization does involve a tradeoff, but cost is small

- Other questions: subgroup analysis, pre-analysis plans, . . .