# Discrepancy Theory and Randomized Controlled Trials


Daniel A. Spielman

Yale

YALE FDS
INSTITUTE FOR FOUNDATIONS OF DATA SCIENCE

Computer Science
Statistics & Data Science
Mathematics


Harvard Oct '23

# BALANCING COVARIATES IN RANDOMIZED EXPERIMENTS USING THE GRAM–SCHMIDT WALK

By Christopher Harshaw, Fredrik Sävje,
Daniel A. Spielman and Peng Zhang (张鹏)

Chris
Harshaw
MIT

Fredrik
Sävje
Yale

Peng
Zhang
Rutgers

Randomized Controlled Trials (RCT)

Test medications, procedures or policies

Randomly assign subjects to drug or placebo

Want the two groups to be similar.

Discrepancy Theory

Divide a group of things (vectors)
into two similar groups.

More similar than random

# Outline

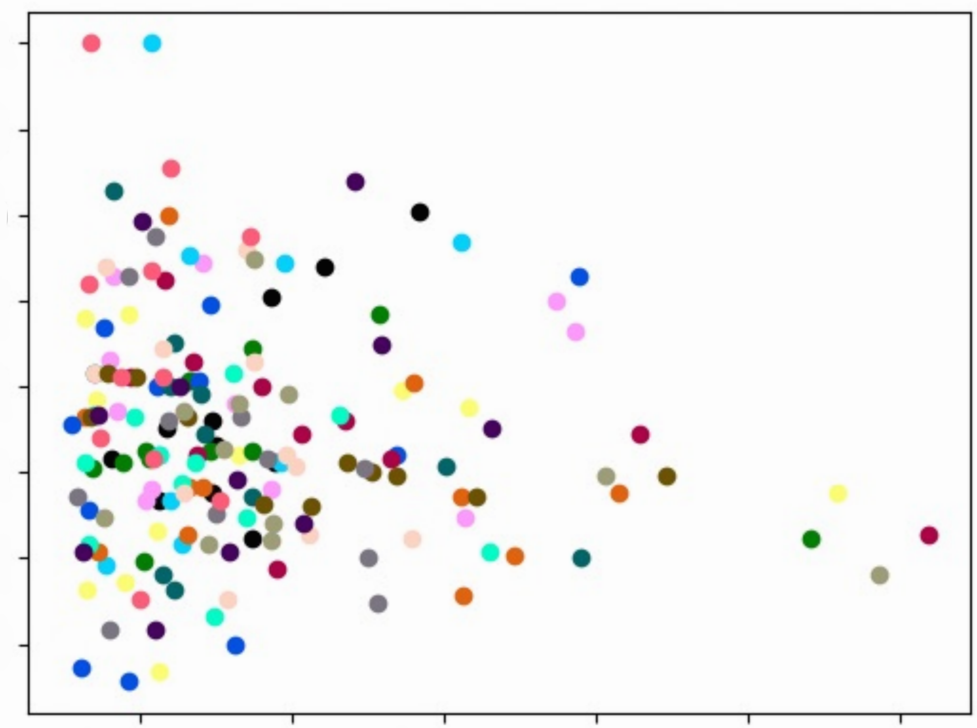
Introduction to Discrepancy Theory

How to analyze RCTs
probability & statistics

Discrepancy theory for RCTs

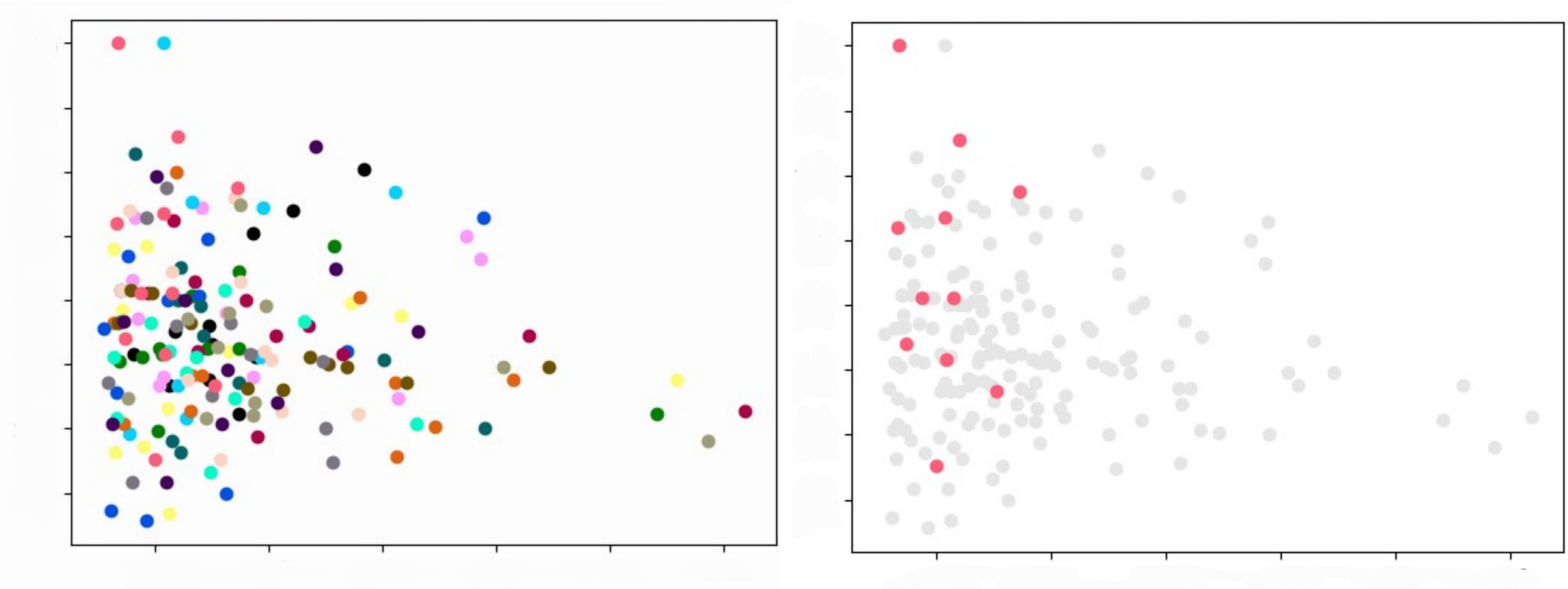
# Village-wide maternal health intervention

176 villages, 16 treatments, 11 villages per group



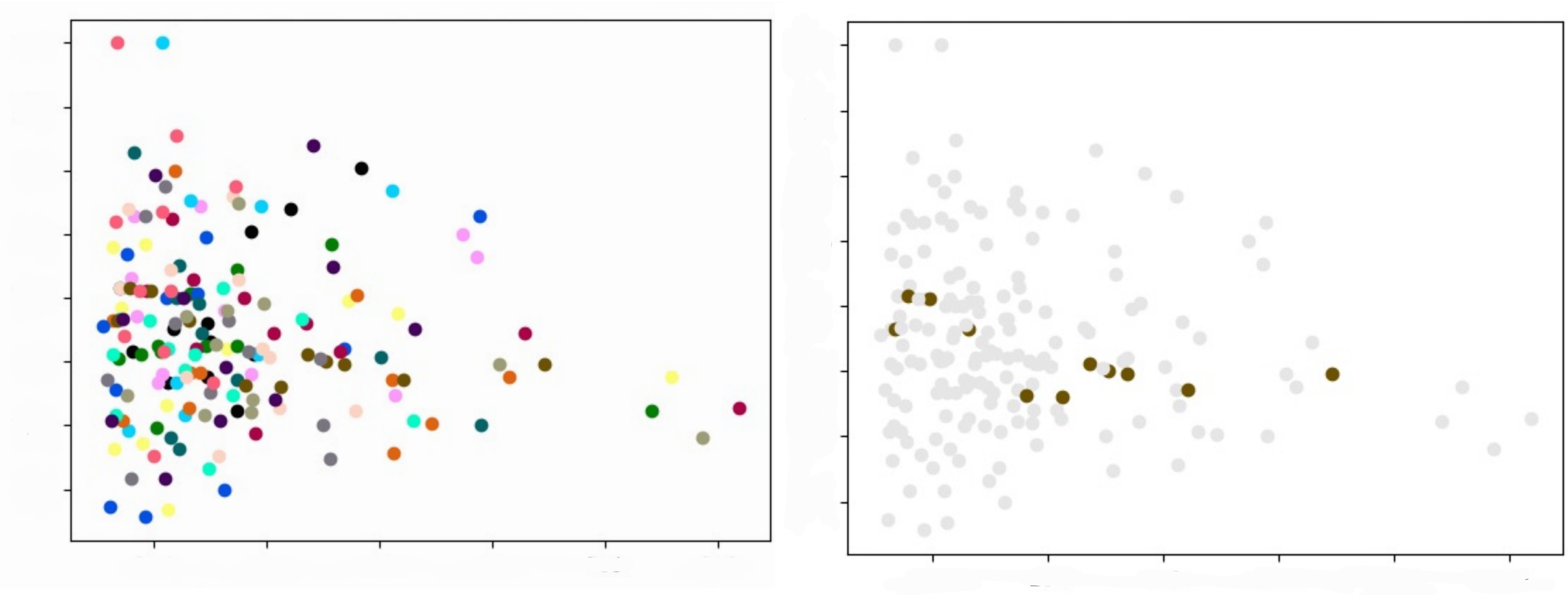
# Village-wide maternal health intervention

176 villages, 16 treatments, 11 villages per group



# Village-wide maternal health intervention

176 villages, 16 treatments, 11 villages per group



Want to balance many **covariates**

Distance to hospital

Time to reach hospital when raining

Median income

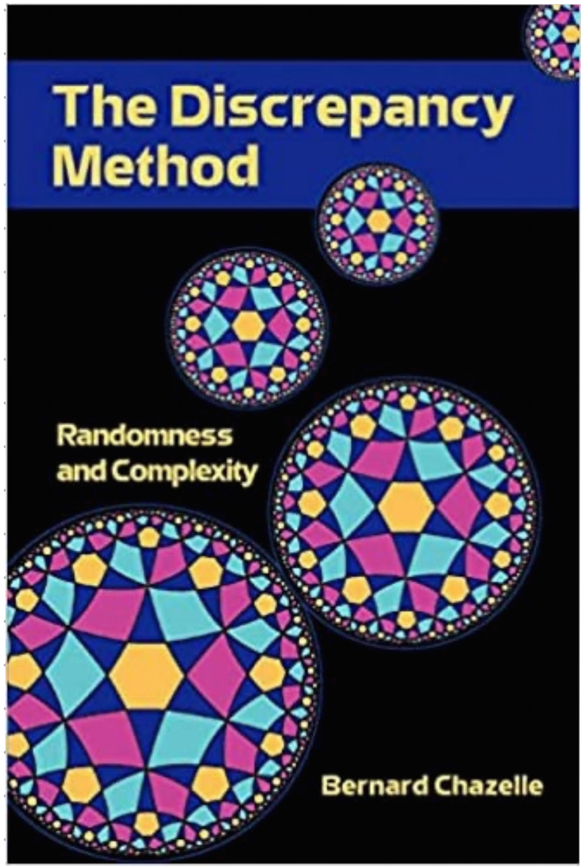
Average age

Altitude

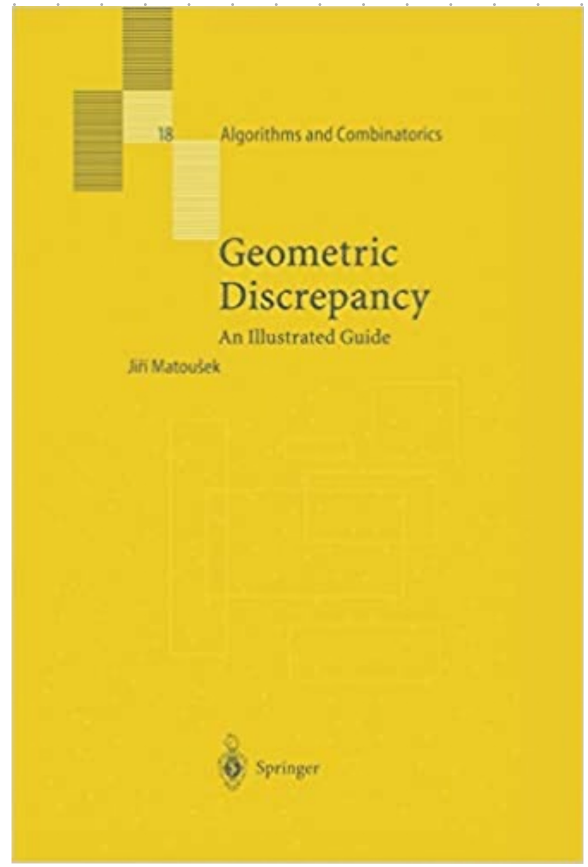
...

# Discrepancy Theory

Chazelle



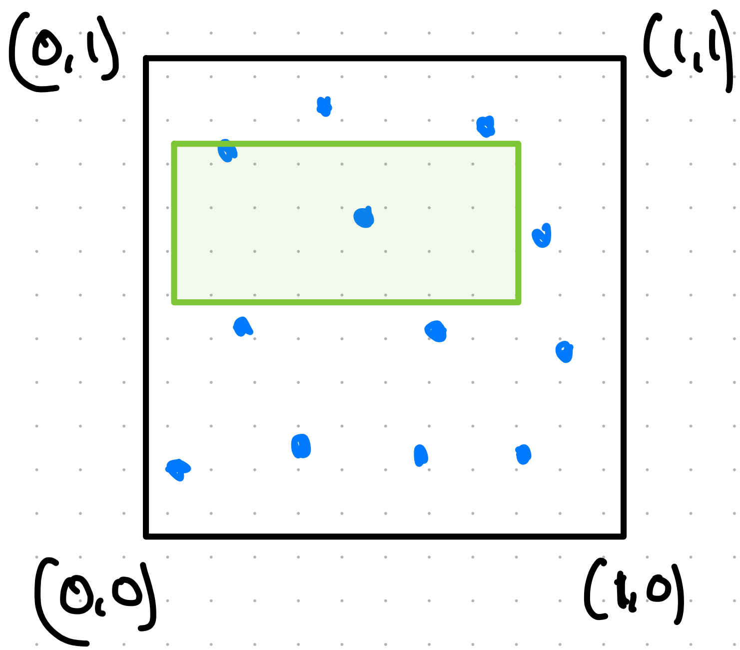

Matoušek




Courses by Aleksandar Nikolov
Peng Zhang

# Discrepancy Theory:

Illustration by points and axis-parallel rectangles in $[0,1]^2$

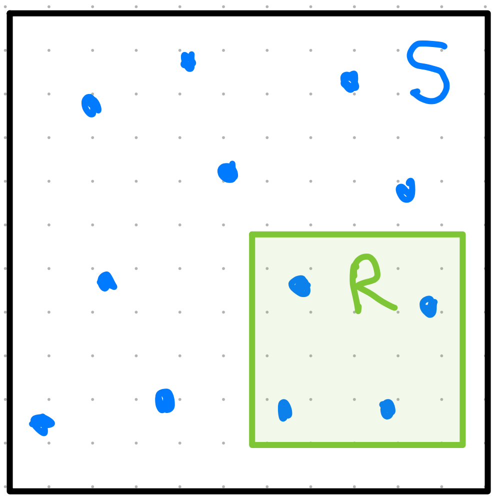


# Discrepancy Theory: axis-parallel rectangles

Problem: find $S = \{x^1, \dots, x^n\} \subseteq [0,1]^2$

so that for all rectangles $R$,

$$\frac{|S \cap R|}{|S|} \approx \text{volume}(R)$$




How good can this approximation be?

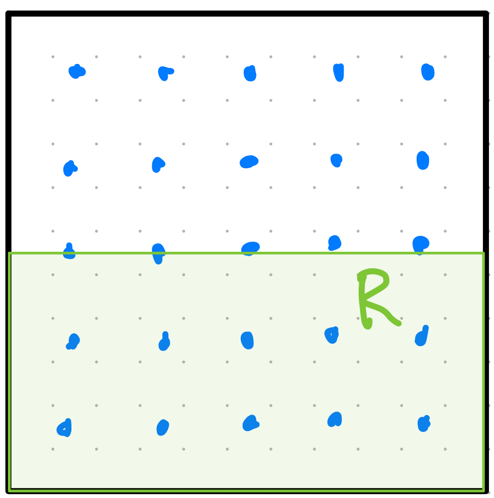
$$\frac{|S \cap R|}{|S|} = \frac{4}{12} = \frac{1}{3}$$

# Discrepancy Theory: axis-parallel rectangles

Given $S = \{x^1, \ldots, x^n\} \subseteq [0,1]^2$ and rectangle $R$

discrepancy is $\left| \dfrac{|S \cap R|}{|S|} - \text{volume}(R) \right|$.

Want small discrepancy for all $R$




$S$ = grid points

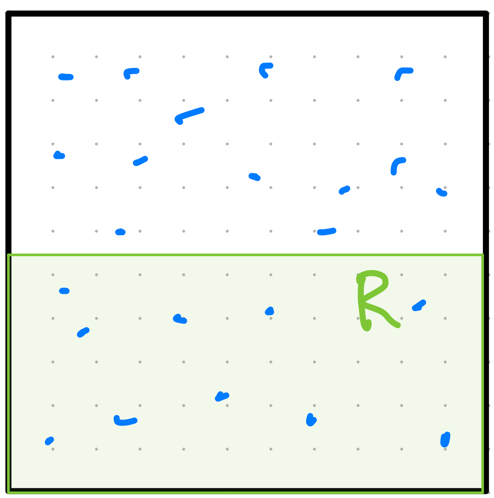
$\exists\, R$ with discrepancy $\geq \dfrac{1}{2\sqrt{n}}$

# Discrepancy Theory: axis-parallel rectangles

Given $S = \{x^1, \ldots, x^n\} \subseteq [0,1]^2$ and rectangle $R$

discrepancy is $\left| \frac{|S \cap R|}{|S|} - \text{volume}(R) \right|$.

Want small discrepancy for all $R$




$S$ = random points

$R = [0,1] \times [0, 1/2]$

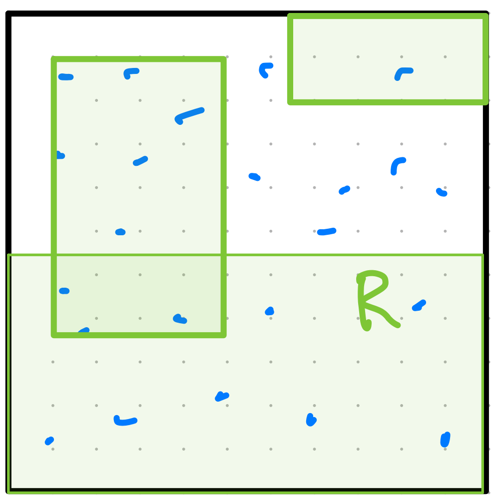
Expected discrepancy $\approx \frac{1}{\sqrt{2\pi n}}$

# Discrepancy Theory: axis-parallel rectangles

Given $n$, find $S = \{x^1, \ldots, x^n\} \subseteq [0,1]^2$ so that $\forall R$

discrepancy $\left| \dfrac{|S \cap R|}{|S|} - \text{volume}(R) \right|$ is small

**Theorem** (Van der Corput '35)




There exists $S$ such that for all $R$,

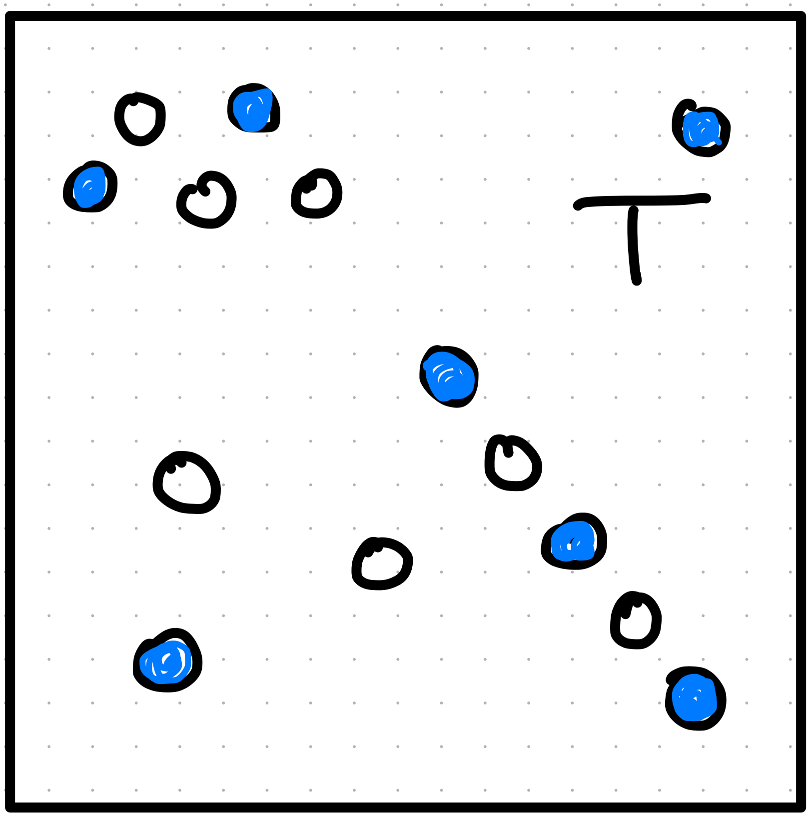
discrepancy $\leq \dfrac{c \log n}{n}$,

for some constant $c$.

# A Partitioning Problem

Given $T \subseteq [0,1]^2$ and $n$, find $S \subseteq T$, $|S| = n$ s.t.

$$\forall R \quad \left| \frac{|S \cap R|}{|S|} - \frac{|T \cap R|}{|T|} \right| \quad \text{is small}$$




usually, $n = \frac{|T|}{2}$

S approximates T

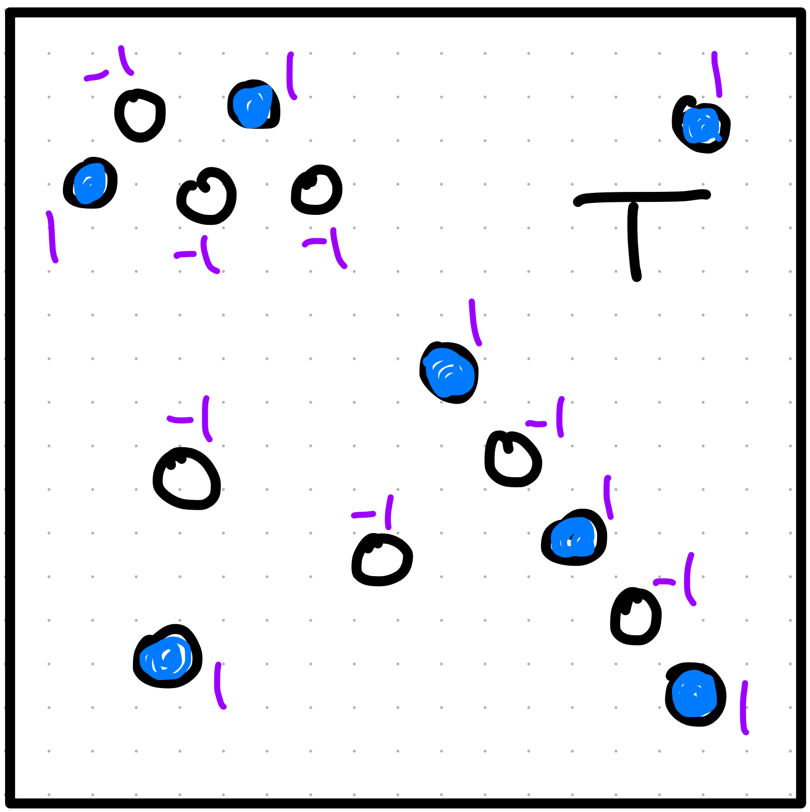
and so does T−S

# A Partitioning Problem

Given $T \subseteq [0,1]^2$ and $n$, find $S \subseteq T$, $|S| = n$ s.t.

$$\forall R \quad \left| \frac{|S \cap R|}{|S|} - \frac{|T \cap R|}{|T|} \right| \quad \text{is small}$$

Let $z(j) = \begin{cases} 1 & \text{if } j \in S \\ -1 & \text{o.w.} \end{cases}$




usually, $n = \frac{|T|}{2}$

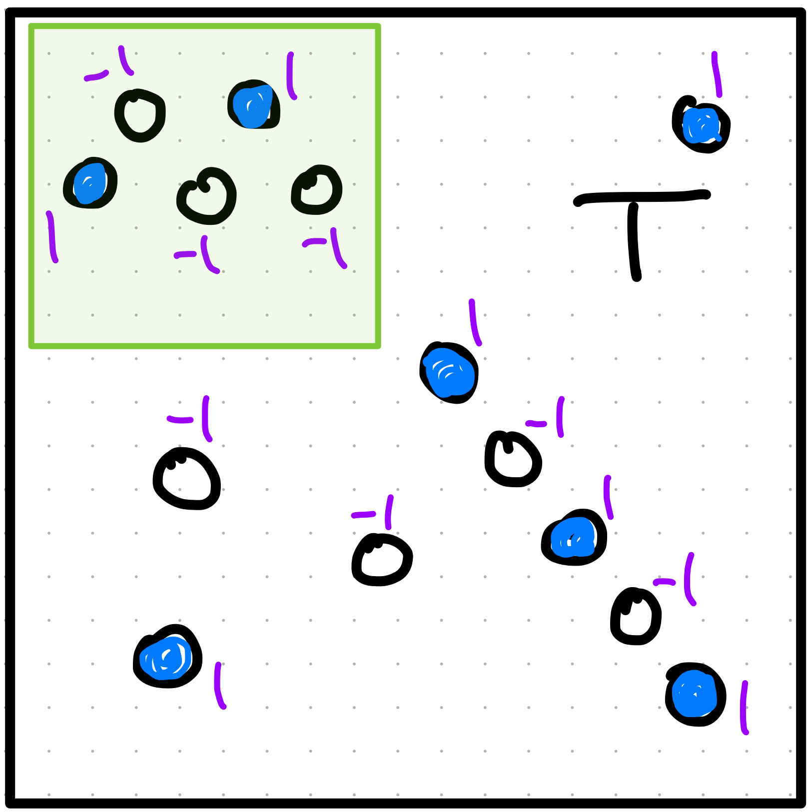
$S$ approximates $T$

and so does $T - S$

# A Partitioning Problem

Given $T \subseteq [0,1]^2$ and $n$, find $z : T \to \{\pm 1\}$ s.t.

$$\forall R \quad \left| \sum_{j \in R} z(j) \right| \quad \text{is small}$$




Each rectangle gives a subset of $T$

# Discrepancy for Set Systems

Define $[n] = \{1, \dots, n\}$

Given $S_1 \subseteq [n]$, $S_2 \subseteq [n]$, $\dots$, $S_d \subseteq [n]$

Find $z : [n] \to \{\pm 1\}$ s.t.

$$\forall i \quad \left| \sum_{j \in S_i} z(j) \right| \text{ is small}$$

Define $\operatorname{disc}(z) = \max_i \left| \sum_{j \in S_i} z(j) \right|$

# Discrepancy for Set Systems

Define $[n] = \{1, \dots, n\}$

Given $S_1 \subseteq [n]$, $S_2 \subseteq [n]$, $\dots$, $S_d \subseteq [n]$

Define $\mathrm{disc}(z) = \max_i \left| \sum_{j \in S_i} z(j) \right|$

For uniform random $z$,

$$\Pr_z \left[ \mathrm{disc}(z) \le \sqrt{2n \ln(2n)} \right] > 0$$

# Discrepancy for Set Systems

Define $[n] = \{1, \dots, n\}$

Given $S_1 \subseteq [n]$, $S_2 \subseteq [n]$, …, $S_n \subseteq [n]$

Define $\mathrm{disc}(z) = \max_i \left| \sum_{j \in S_i} z(j) \right|$

For uniform random $z$,

$$\Pr_z\left[ \mathrm{disc}(z) \leq \sqrt{2n \ln(2d)} \right] > 0$$

## Spencer '85

$\exists z$ s.t. $\mathrm{disc}(z) \leq 6\sqrt{d}$ when $d \leq n$

# Discrepancy for Set Systems → Vectors

Given $S_1 \subseteq [n]$, $S_2 \subseteq [n]$, …, $S_d \subseteq [n]$

let $X$ be the incidence matrix of the sets

$$X(i,j) = \begin{cases} 1 & \text{if } j \in S_i \\ 0 & \text{o.w.} \end{cases} \qquad (Xz)(i) = \sum_{j \in S_i} z(j)$$

$$\text{disc}(z) = \max_i \Big| \sum_{j \in S_i} z(j) \Big| = \|Xz\|_\infty$$

where $\|v\|_\infty \stackrel{\text{def}}{=} \max_j |v(j)|$

Let $X = \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{pmatrix}$ so $Xz = \sum_j z(j)\, x_j$

# Discrepancy for Vectors

**Spencer '85**

Given $x_1 \in [-1,1]^d, \ldots, x_n \in [-1,1]^d$, $X = \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{pmatrix}$

$\exists z \in \{\pm 1\}^n$ s.t. $\|Xz\|_\infty \leq 6\sqrt{d}$, $d \leq n$

Was unknown if could find $z$
in polynomial time.

# Discrepancy for Vectors

**Spencer '85**

Given $x_1 \in [-1,1]^d, \ldots, x_n \in [-1,1]^d$, $X = \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{pmatrix}$

$\exists z \in \{\pm 1\}^n$ s.t. $\|Xz\|_\infty \leq 6\sqrt{d}$

**Bansal '10** Can find $z$ in polynomial time.

Followed by many other algorithms
for many discrepancy problems

We still do not know the right constant. Is it 2?

# Euclidean Discrepancy for Vectors

Bárány - Grinberg '81 + Beck - Fiala '81

Given $x_1, \dots, x_n \in \mathbb{R}^d$, $\|x_i\|_2 \leq 1$ for all $i$

$\exists z \in \{\pm 1\}^n$ s.t. $\|\sum_i z(i) x_i\|_2 = \|Xz\|_2 \leq \sqrt{d}$

Cannot improve $\sqrt{d}$ : $x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ $x_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ $x_d = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

$x_i = 0 \quad i > d$

# Euclidean Discrepancy for Vectors

Bárány - Grinberg '81 + Beck - Fiala '81 (algorithmic)

**Thm 1** Given $X_1, \dots, X_n \in \mathbb{R}^d$, $\|X_i\|_2 \leq 1$ for all $i$

$$\exists z \in \{\pm 1\}^n \text{ s.t. } \|Xz\|_2 \leq \sqrt{d}$$

---

Charikar, Newman and Nikolov '11:

Is NP-hard to distinguish

$$\exists z \in \{\pm 1\}^n \text{ s.t. } Xz = \bar{0}$$

from

$$\forall z \in \{\pm 1\}^n \text{ s.t. } \|Xz\|_2 \geq c\sqrt{d}$$

# Banaszczyk's '98 Theorem

Generalizes Euclidean discrepancy

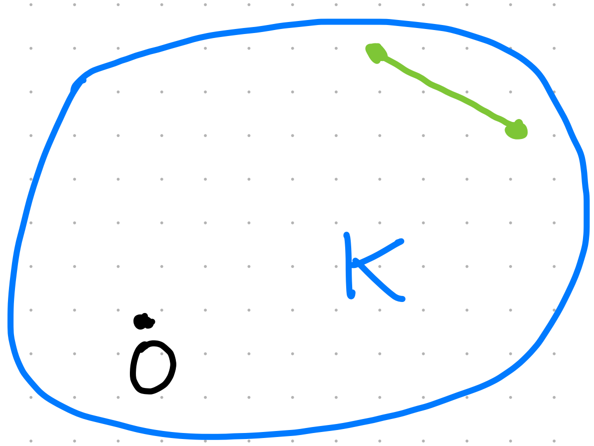
Given $x_1, \dots, x_n \in \mathbb{R}^d$, $\|x_i\|_2 \le 1$ for all $i$

Thm 1 says $\exists z \in \{\pm 1\}^n$ s.t. $\sum_i z(i) x_i \in B_{\sqrt{d}}(0)$

Banaszczyk:

For all convex $K$ with Gaussian measure $\ge \frac{1}{2}$

$\exists z \in \{\pm 1\}^n$ s.t. $\sum_i z(i) x_i \in 5K$



# Banaszczyk's '98 Theorem

Generalizes Euclidean discrepancy

Given $x_1, \dots, x_n \in \mathbb{R}^d$, $\|x_i\|_2 \le 1$ for all $i$

Thm 1 says $\exists z \in \{\pm 1\}^n$ s.t. $\sum_i z(i) x_i \in B_{\sqrt{d}}(0)$

Banaszczyk:

For all convex $K$ with Gaussian measure $\ge \frac{1}{2}$

$$\exists z \in \{\pm 1\}^n \text{ s.t. } \sum_i z(i) x_i \in 5K$$

$$\text{Gaussian measure} = \Pr_{x \leftarrow N(0, I_d)}[x \in K] = \frac{1}{(2\pi)^{d/2}} \int_{x \in K} e^{-\frac{1}{2}\|x\|_2^2} dx$$

# Banaszczyk's '98 Theorem

Given $x_1, \dots, x_n \in \mathbb{R}^d$, $\|x_i\|_2 \le 1$ for all $i$

Thm 1 says $\exists z \in \{\pm 1\}^n$ s.t. $\sum_i z(i) x_i \in B_{\sqrt{d}}(0)$

Banaszczyk:

For all convex $K$ with Gaussian measure $\geq \frac{1}{2}$

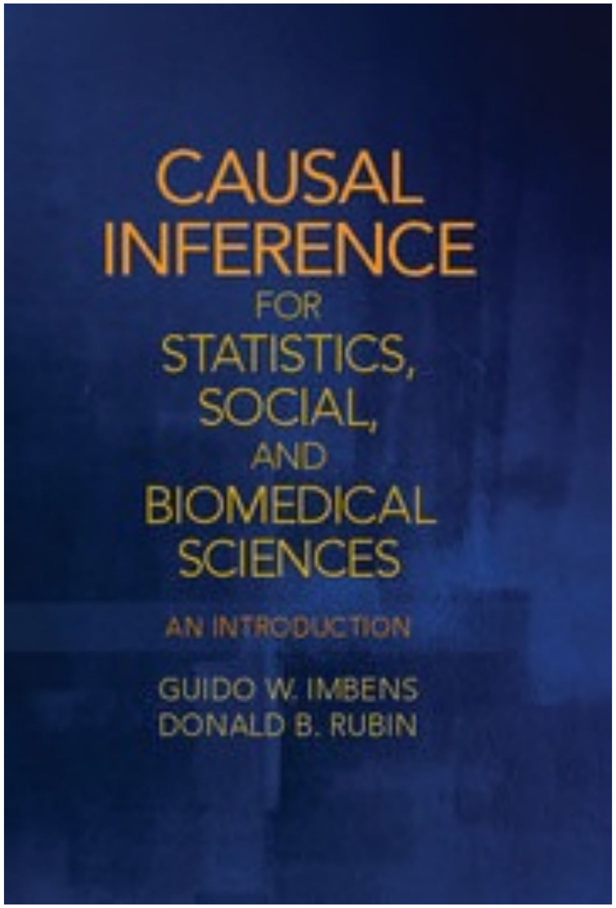
$\exists z \in \{\pm 1\}^n$ s.t. $\sum_i z(i) x_i \in 5K$

Efficient algorithm, Gram-Schmidt Walk

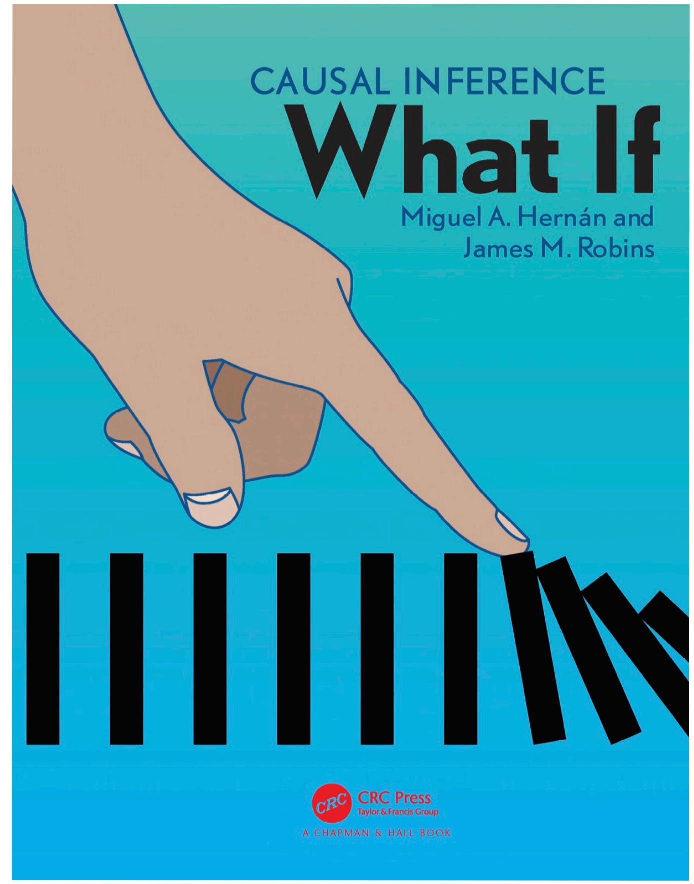
by Dadush, Garg, Lovett, Nikolov '16 +

Bansal, Dadush, Garg, Lovett '19

# Randomized Controlled Trials



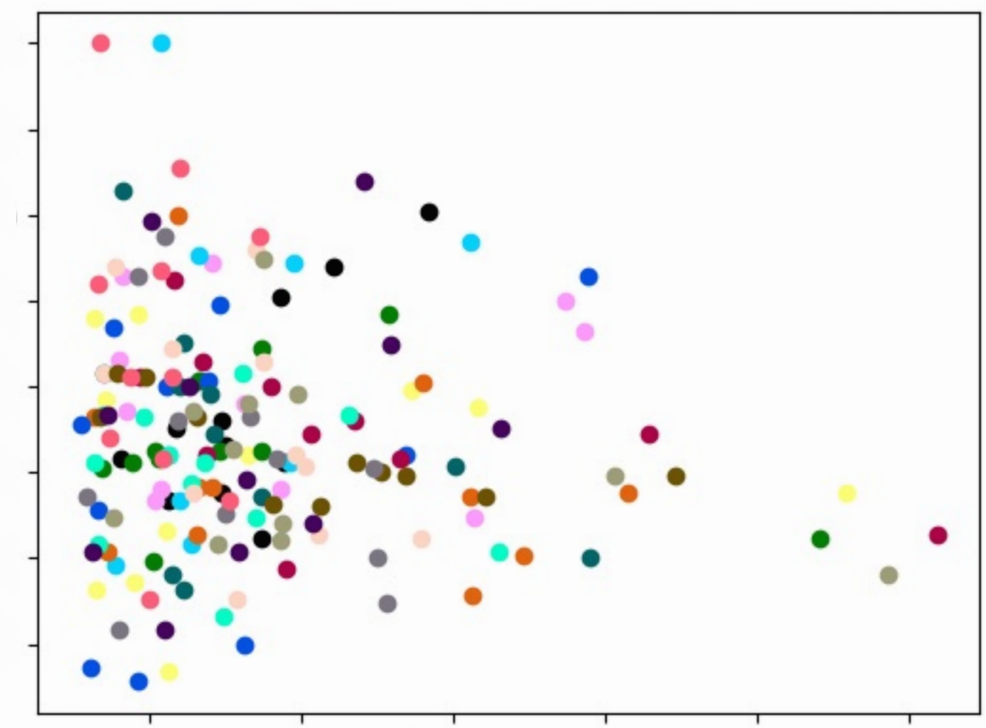

Imbens & Rubin



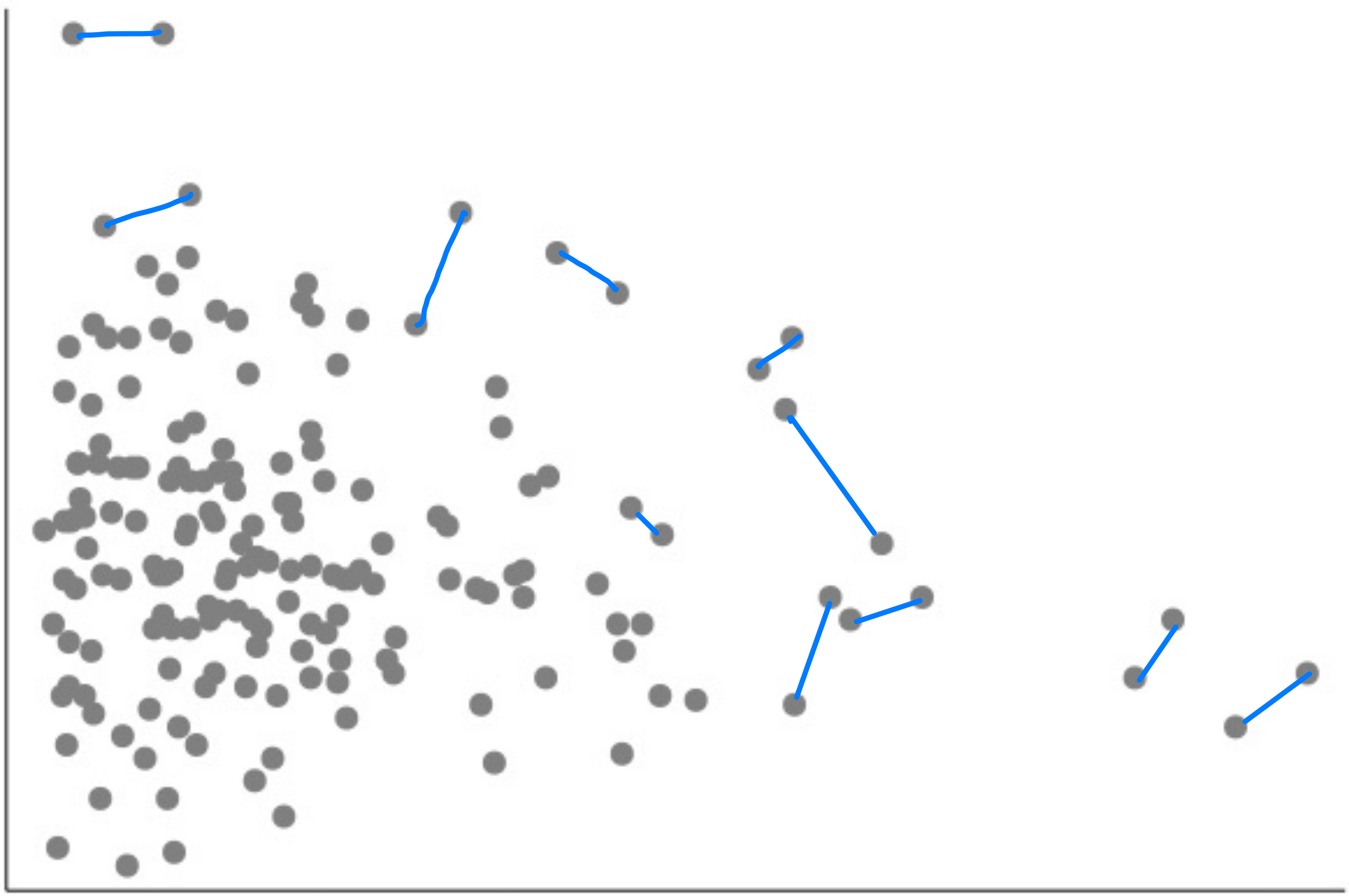

Hernán & Robins

# Village-wide maternal health intervention

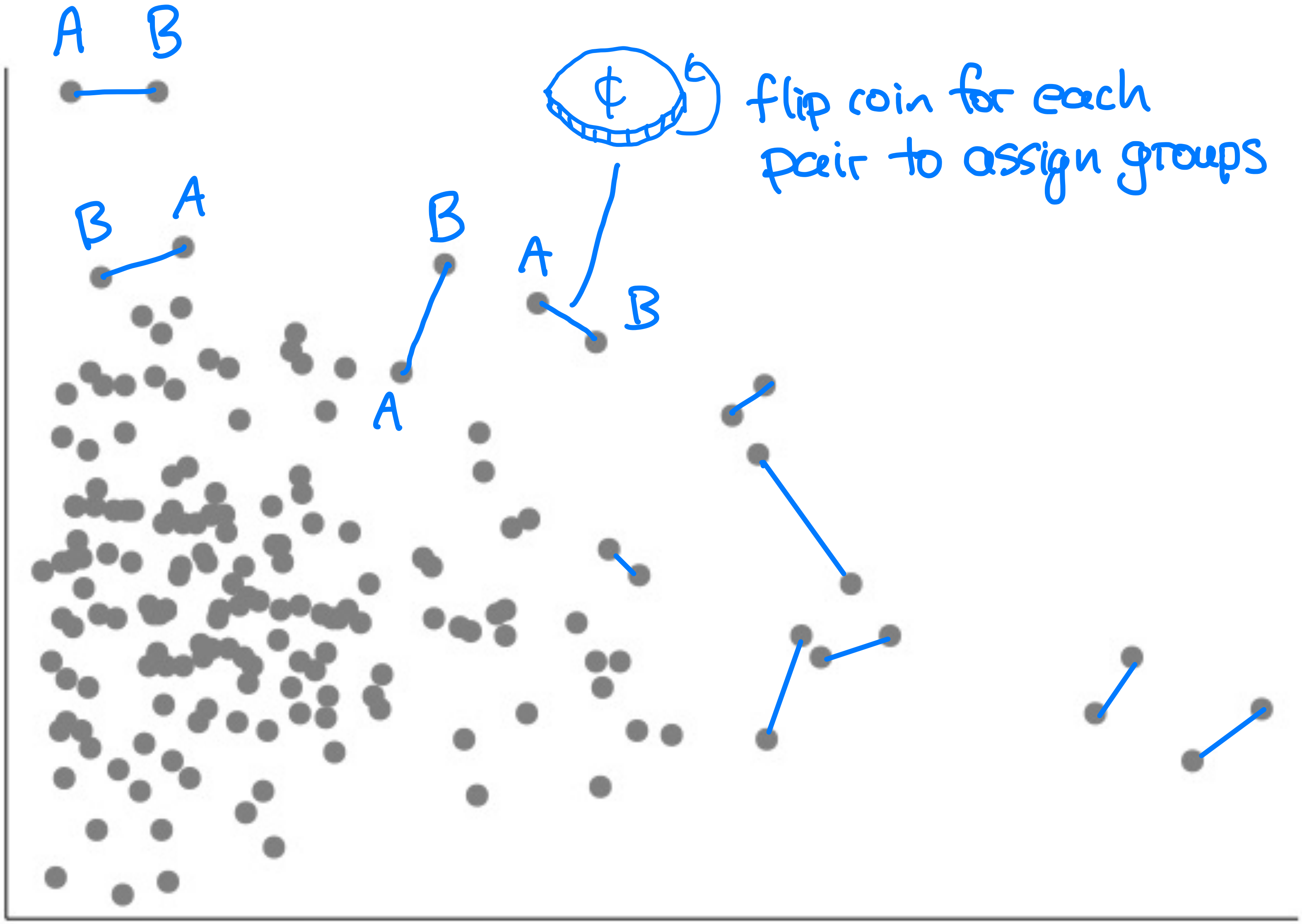
176 villages, 16 treatments, 11 villages per group



# Matching for two groups

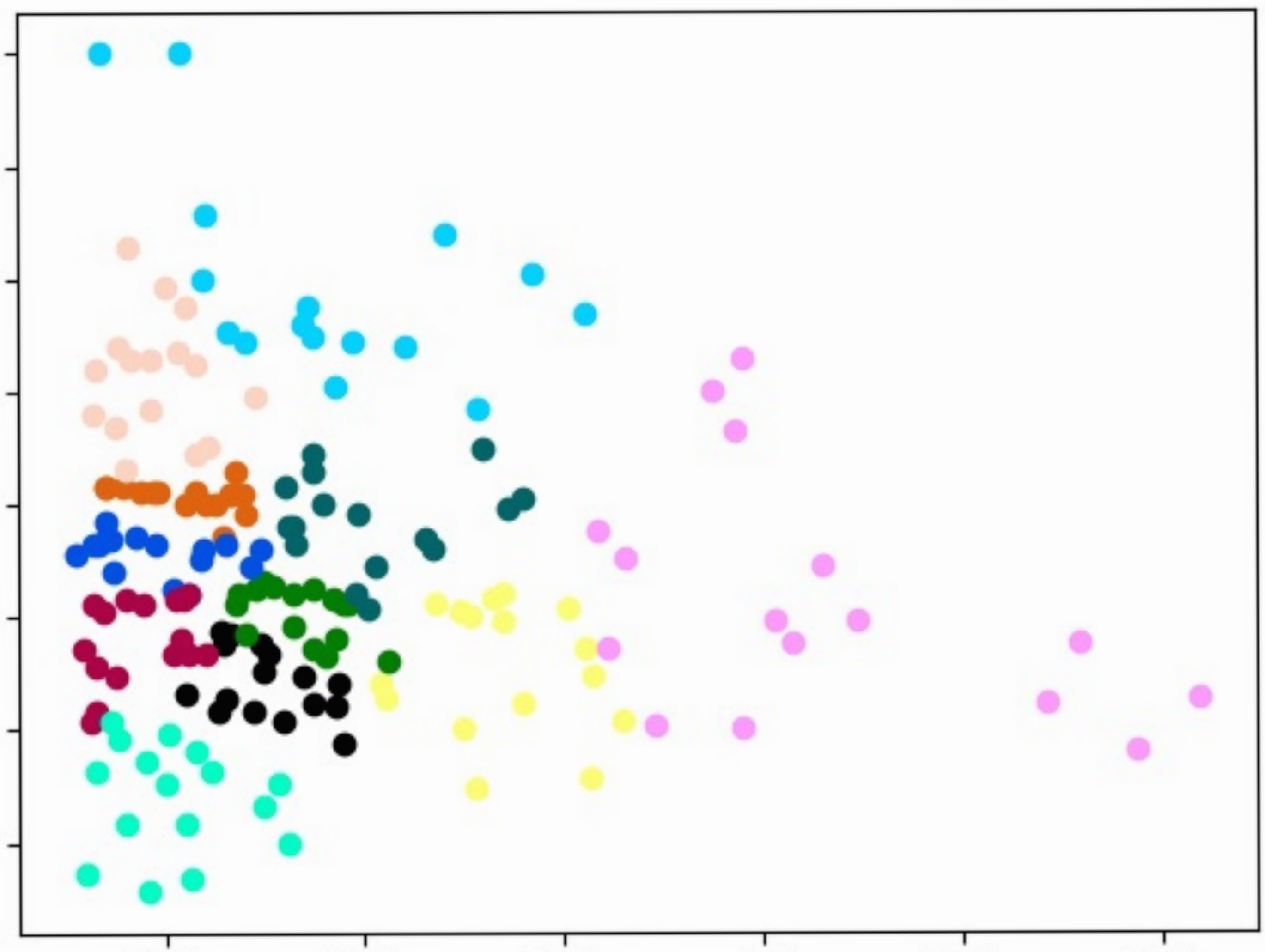


# Matching for two groups



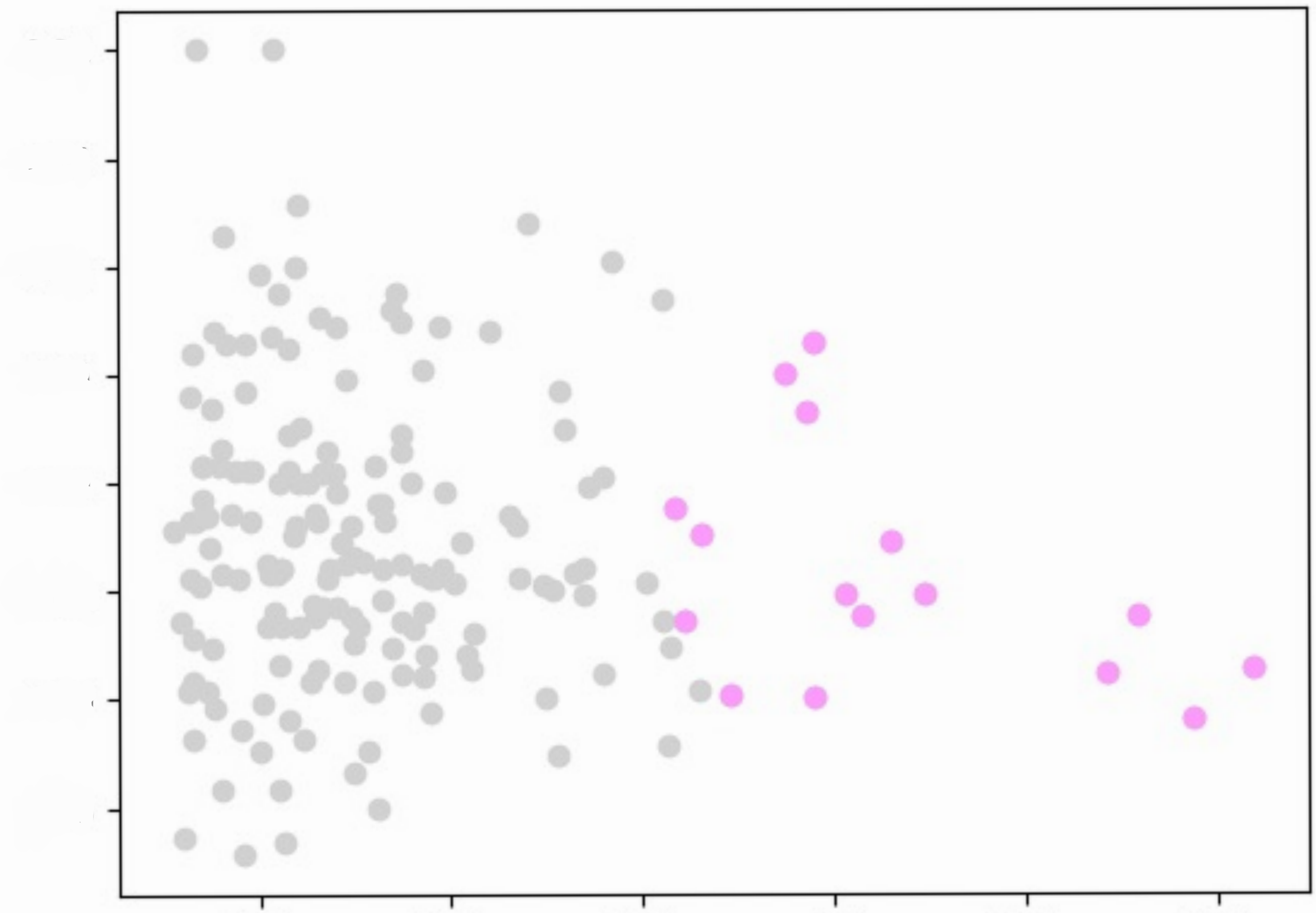
# Blocking for many treatment groups

Create 11 clusters of 16 villages each



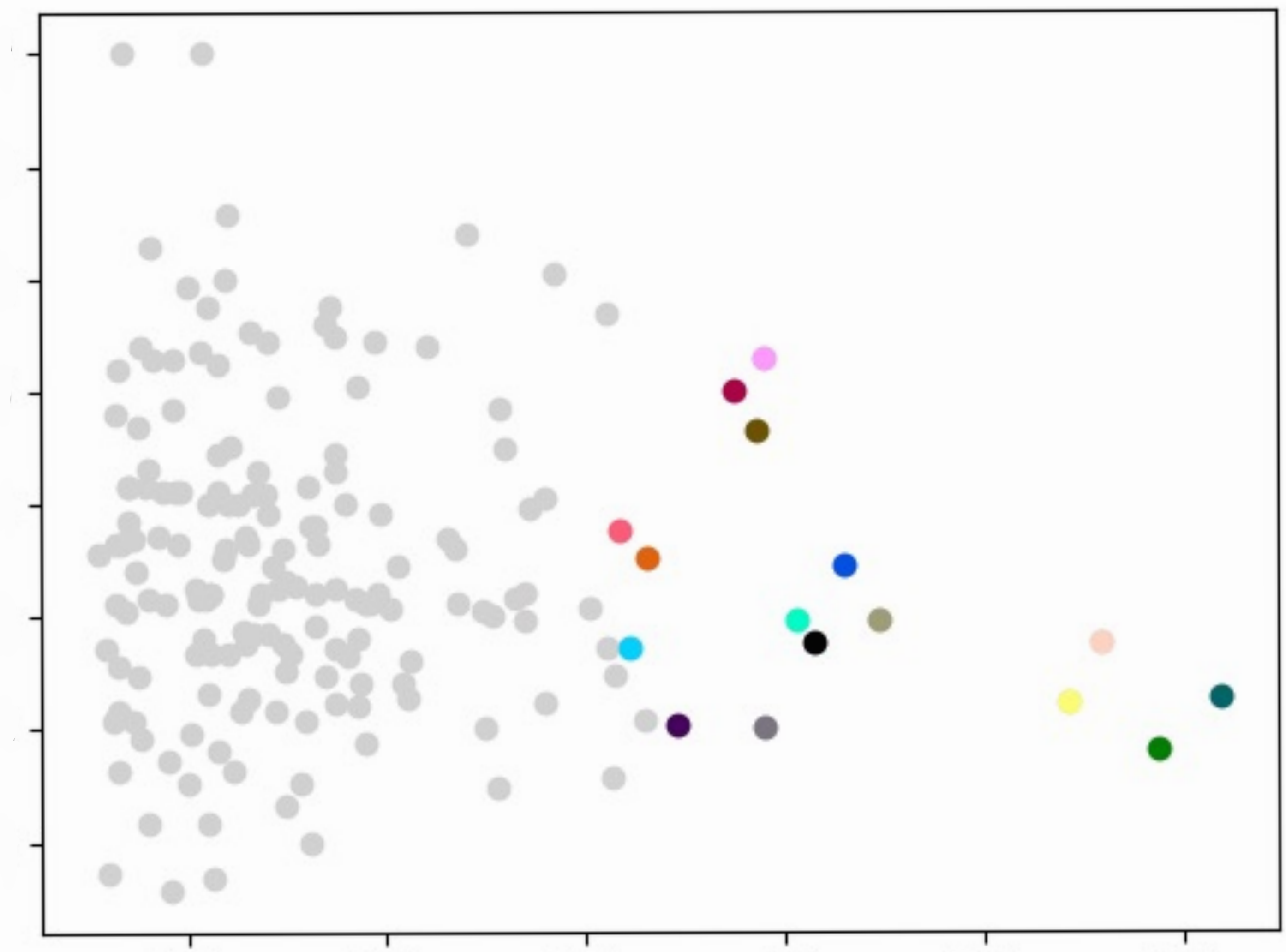
# Blocking for many treatment groups

For each cluster ...



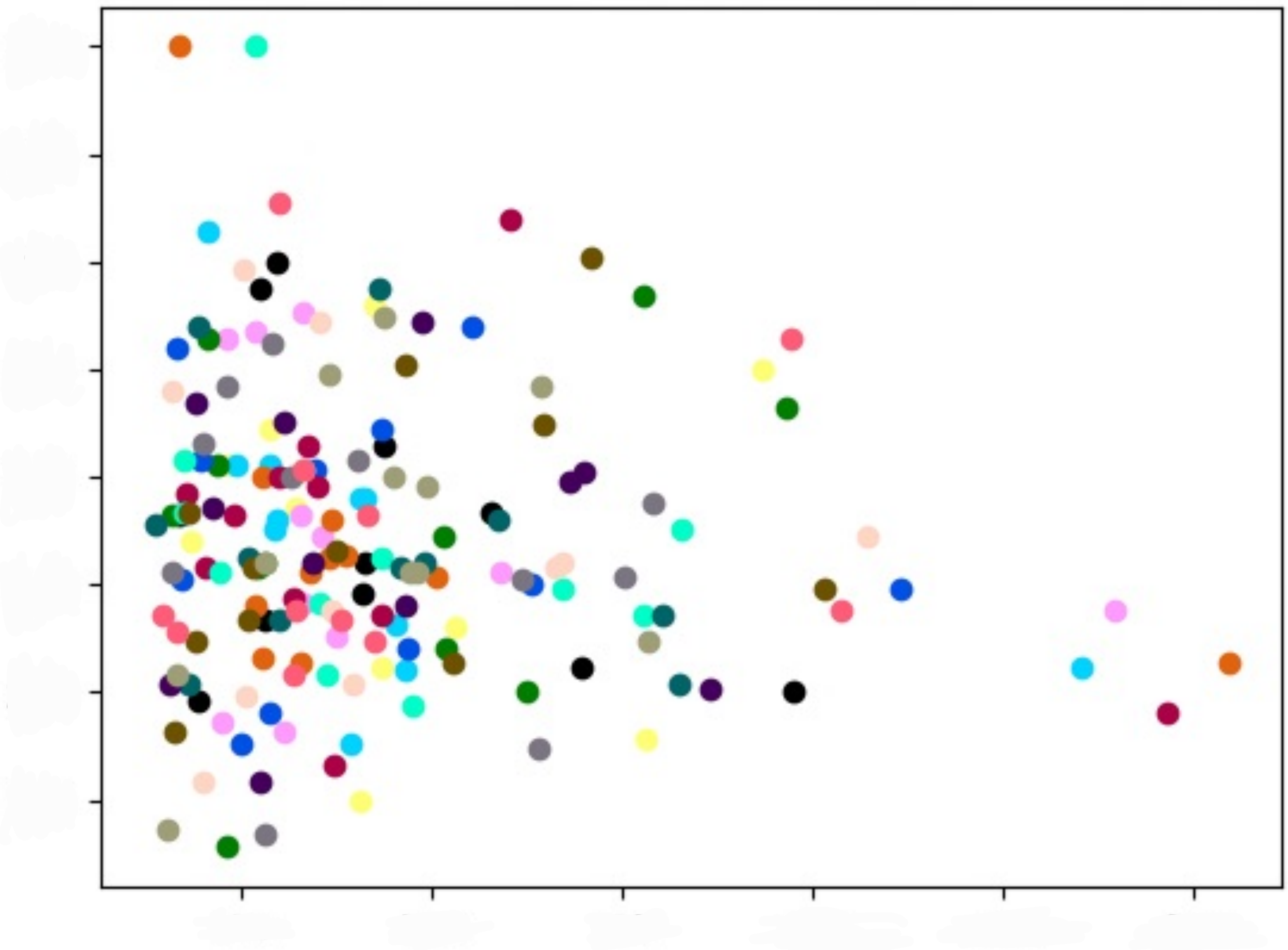
# Blocking for many treatment groups

Choose a random bijection from villages to treatments



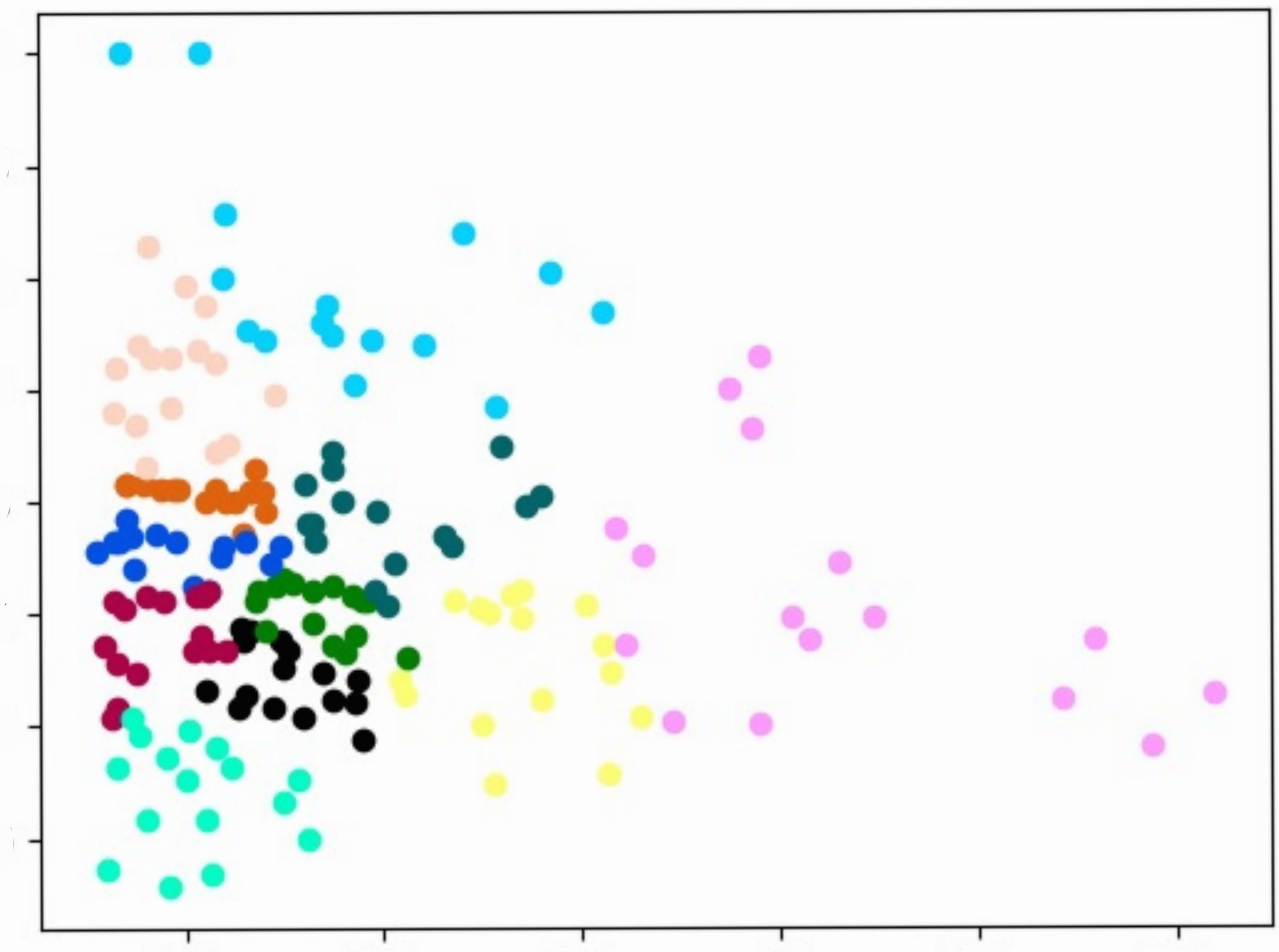
# Blocking for many treatment groups

Assign each cluster independently
to get 16 groups of 11 villages each

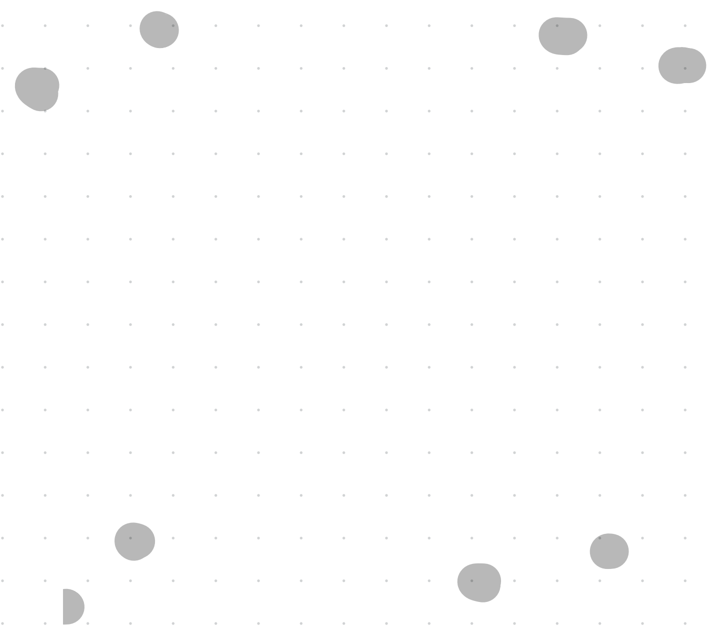


# Blocking for many treatment groups

Create 11 clusters of 16 villages each

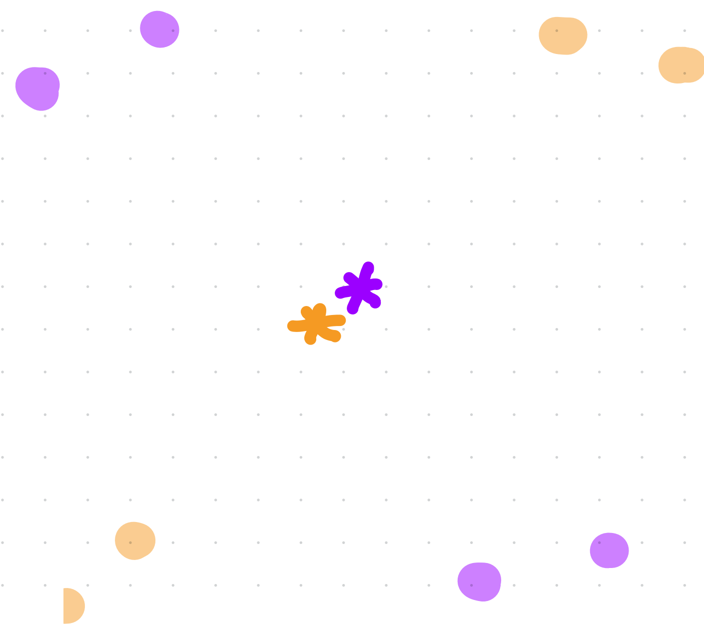


# Discrepancy Theory



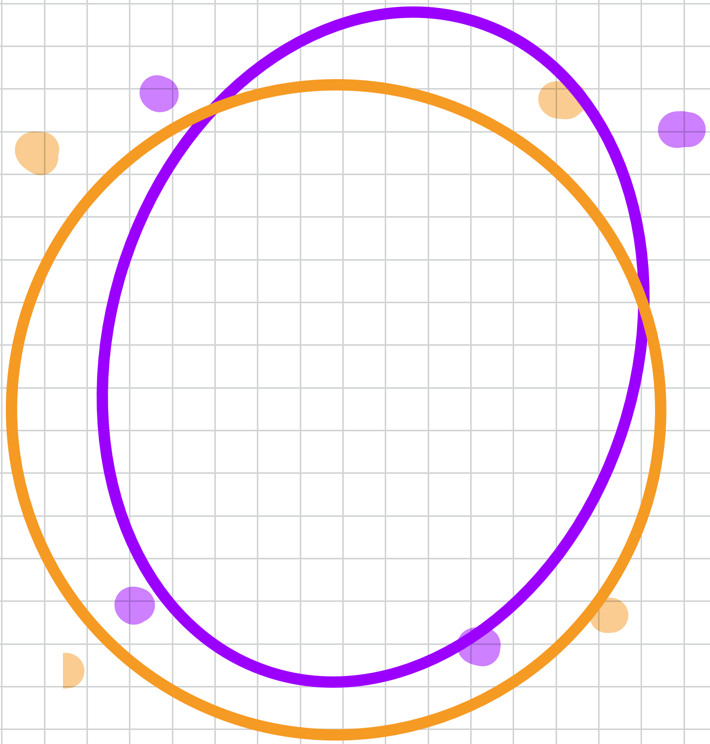
Can make means similar

# Discrepancy Theory



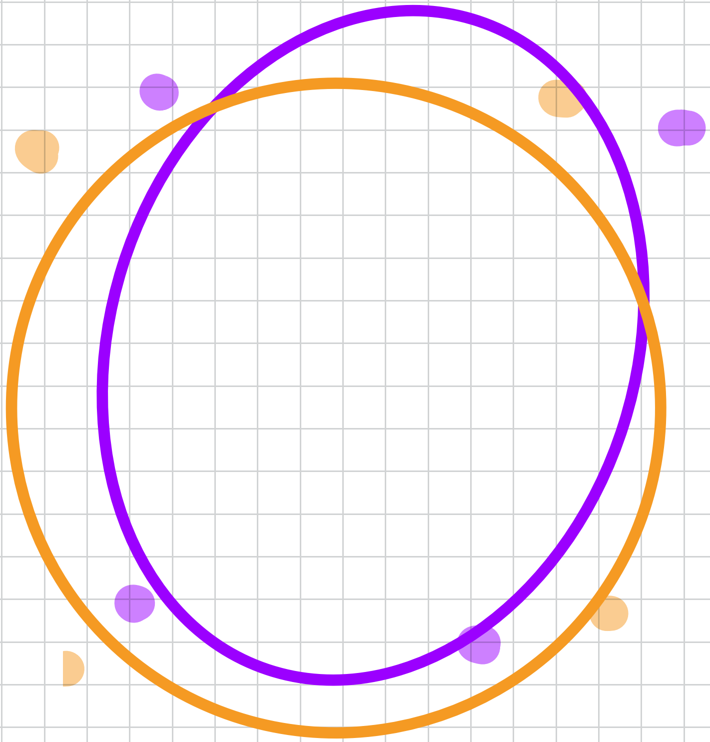
Can make means similar

Similar moment ellipses always possible



Marcus-S-Srivastava → Kadison-Singer → Weaver

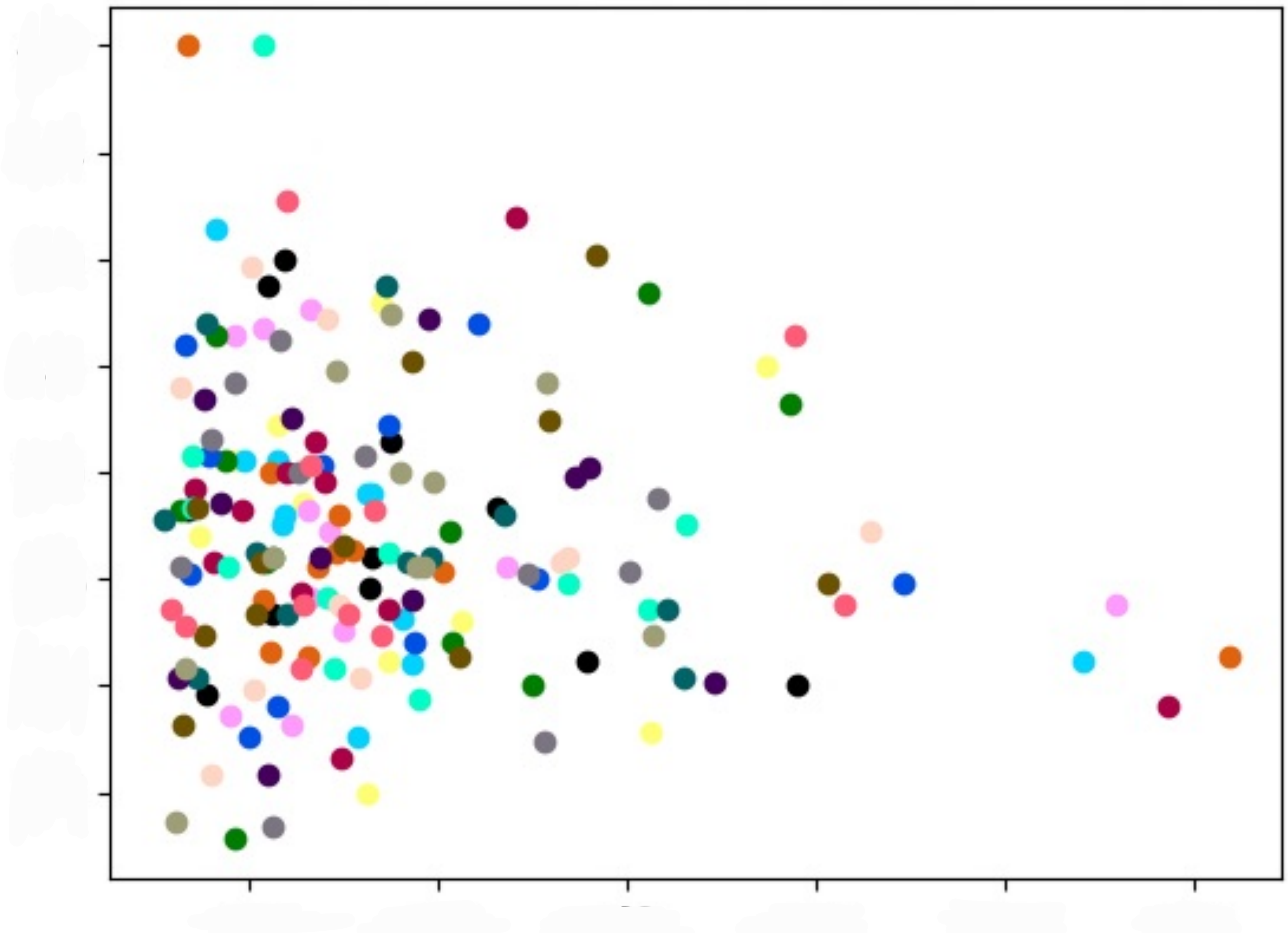
Similar moment ellipses always possible



Marcus-S-Srivastava → Kadison-Singer → Weaver

Open: Balance means and second moments

# Balanced Everything



# Randomization

Allows one to argue results unlikely to be due to a poor assignment.

Need randomization for confidence intervals

Can balance better on average

This is not an optimization problem!

100 balanced but very different assignments

1. Framework & Mathematical Formalization

2. Algorithm

The Gram-Schmidt Walk Design is

**Robust**: never much worse than iid random

**Balanced**: much better when covariates correlate with outcomes

# Neyman-Rubin Model of Experiments

Consider two treatment groups:

A = Test and B = Control

Two potential outcomes for subject $i$

observe $a(i)$ if $i \in A$

$b(i)$ if $i \in B$

Can not observe both

Assignment is only source of randomness

Response to treatment is <u>NOT</u> random

# Designs

The design is the distribution of A, B

uniform: $\Pr[i \in A] = 1/2$, independently

balanced: A uniform subset of half

Matching: divide subjects into disjoint pairs $(i,j)$

assign $i \to A$, $j \to B$ or

$i \to B$, $j \to A$

with prob $1/2$

# Neyman-Rubin Model for Experiments

Two potential outcomes for subject $i$

observe $a(i)$ if $i \in A$ prob $1/2$

$b(i)$ if $i \in B$ prob $1/2$

Can not observe both

Want to measure average treatment effect

$$\tau = \frac{1}{n} \sum_i a(i) - b(i) \qquad \text{(ATE)}$$

# Horvitz-Thompson Estimator

Estimate $\tau = \frac{1}{n} \sum_i a(i) - b(i)$

by $$\hat{\tau} = \frac{2}{n} \left( \sum_{i \in A} a(i) - \sum_{i \in B} b(i) \right)$$

$$\mathbb{E}\, \hat{\tau} = \tau$$

Are not using difference of means

$$\frac{1}{|A|} \sum_{i \in A} a(i) - \frac{1}{|B|} \sum_{i \in B} b(i)$$

Differs from Horvitz-Thompson when $|A| \neq |B|$

# Precision

$$\Pr\left[\,|\hat{\tau} - \tau| > t\,\right]$$

# Confidence Interval

$$\text{Estimator of } \Pr\left[\,|\hat{\tau} - \tau| > t\,\right]$$

# Formula for error of estimator

Set $z(i) = \begin{cases} 1 & i \in A \\ -1 & i \in B \end{cases}$

$$n(\hat{\tau} - \tau) = z^T (a+b) = z^T \mu \text{ where } \mu \overset{\text{def}}{=} a+b$$

proof

$$n\tau = \sum_i a(i) - b(i)$$

$$n\hat{\tau} = 2\sum_{i \in A} a(i) - 2\sum_{i \in B} b(i)$$

$$n(\hat{\tau} - \tau) = \sum_{i \in A} (a(i) + b(i)) - \sum_{i \in B} (a(i) + b(i))$$

How show $\Pr[|\hat{\tau} - \tau| > t]$ is small?

1. Estimate MSE $\stackrel{\text{def}}{=} \mathbb{E}(\hat{\tau} - \tau)^2$

Chebyshev: $\Pr\left[|\hat{\tau} - \tau| > c\sqrt{MSE}\right] \leq \frac{1}{c^2}$

2. Higher moments and

Sub-Gaussian Concentration

# Mean Squared Error

$$MSE(\hat{\tau}) = \frac{1}{n^2} \mathop{\mathbb{E}}_{z} (z^T \mu)^2 \qquad z(i) = \begin{cases} 1 & i \in A \\ -1 & i \in B \end{cases}$$

$$= \frac{1}{n^2} \mathop{\mathbb{E}}_{z} (\mu^T z)(z^T \mu)$$

$$= \frac{1}{n^2} \mu^T \left( \mathop{\mathbb{E}}_{z} z z^T \right) \mu$$

$$= \frac{1}{n^2} \mu^T Z \mu$$

$Z \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{z} z z^T$ is the covariance matrix of $z$

# Mean Squared Error

$$\mathrm{MSE}(\hat{\tau}) = \frac{1}{n^2} \mathbb{E}_z (z^T \mu)^2 \qquad z(i) = \begin{cases} 1 & i \in A \\ -1 & i \in B \end{cases}$$

$$= \frac{1}{n^2} \mathbb{E}_z (\mu^T z)(z^T \mu)$$

$$= \frac{1}{n^2} \mu^T (\mathbb{E}_z z z^T) \mu$$

$$= \frac{1}{n^2} \mu^T Z \mu$$

$$\leq \frac{1}{n^2} \|\mu\|^2 \|Z\|_2$$

operator norm of $Z$

We can know $Z$, but not $\mu$

So, want $\|Z\|_2$ small

# Measure Robustness by $\|Z\|_2$

There is a $\mu$ such that

$$\mu^T Z \mu = \|\mu\|^2 \|Z\|_2$$

So, worst-case mean squared error is captured by $\|Z\|_2$

# Robustness - iid case

For $z$ uniform in $\{\pm 1\}^n$

$Z = Cov(z) = I$ and $\|Z\| = 1$

and $MSE = \frac{1}{n^2}\|\mu\|^2$

# Uniform Design Maximizes Robustness

For $z$ uniform in $\{\pm 1\}^n$

$Z = \mathrm{Cov}(z) = I$ and $\|Z\| = 1$

For every other design, $\|Z\| \geq 1$

To get lower variance,

give up a little in worst case

need nice $\mu$ or helpful covariates

# MSE - Random Balanced, $|A| = |B|$

For uniform $z$ s.t. $z^T \mathbb{1} = 0$

$$Z = \frac{n}{n-1}\left(I - \frac{1}{n}\mathbb{1}\mathbb{1}^T\right) \qquad \|Z\|_2 = \frac{n}{n-1}$$

# MSE - Random Balanced, $|A| = |B|$

For uniform $z$ s.t. $z^T \mathbb{1} = 0$

$$Z = \frac{n}{n-1}\left(I - \frac{1}{n}\mathbb{1}\mathbb{1}^T\right) \qquad \|Z\|_2 = \frac{n}{n-1}$$

$$MSE = \frac{1}{(n-1)^2}\left\|\mu - \frac{1}{n}\mathbb{1}^T\mu\right\|_2^2$$

larger than $\frac{1}{n^2}$

small if mean of $\mu$ far from 0

Can be worse than iid,

but often preferred, especially if $a, b \geq 0$

# Covariates

$x_1, \ldots, x_n \in \mathbb{R}^d$ $\qquad X = \begin{pmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{pmatrix}$

Can help if
$\mu$ close to row span of $X$

Ideal case : $\mu = X^T \beta$

Gives $n^2 MSE(\hat{\tau}) = \mu^T Z \mu = \beta^T X Z X^T \beta$

$$\leq \|\beta\|_2^2 \; \|X Z X^T\|_2$$

As don't know $\beta$, will try to make

$\|X Z X^T\|_2$ small

# Impact of Balance and Robustness

$$MSE(\hat{\tau}) \leq \frac{1}{n^2} \min_{\beta \in \mathbb{R}^d} \Big[ \sigma_z \underbrace{\|\mu - X^T\beta\|_2^2}_{\text{error of linear fit}} + \sigma_x \underbrace{\|\beta\|_2^2}_{\text{complexity of } \beta} + \underbrace{\text{cross term}}_{\text{safe to ignore}} \Big]$$

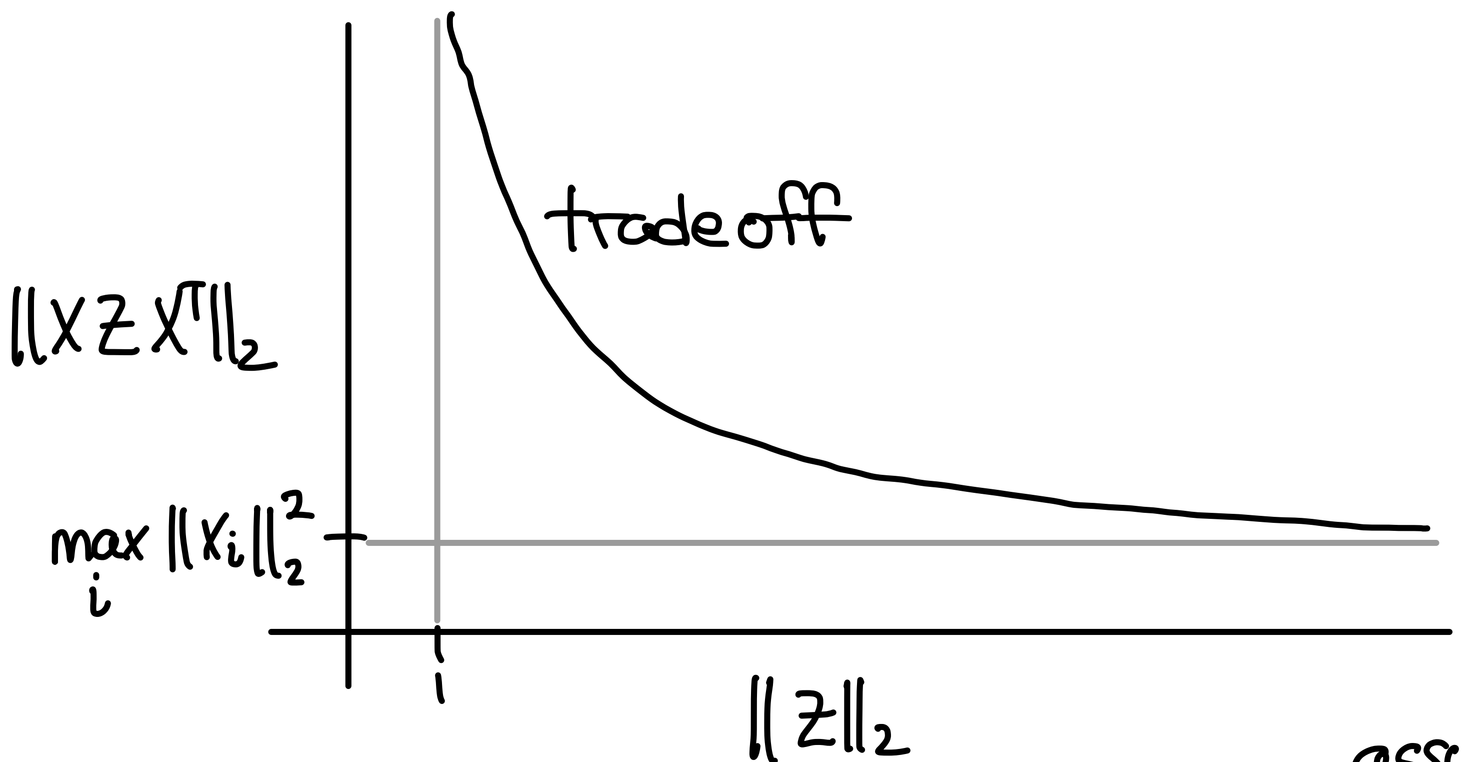
$\sigma_z = \|Z\|_2$ measures Robustness

$\sigma_x = \|XZX^T\|_2$ measures Balance

# Balance - Robustness Tradeoff

| Robustness | Balance |
| --- | --- |
| $\|Z\|_2 \geq 1$ | $\|XZX^T\|_2 \geq 1$ |

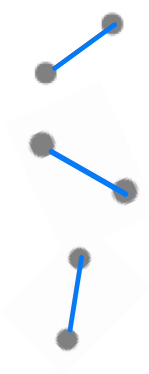
in general




assuming $\|x_i\|_2 \leq 1$

# Matching Design

$$\|Z\|_2 = 2 \qquad Z = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & \\ 0 & 0 & 0 & 0 & & \end{pmatrix}$$



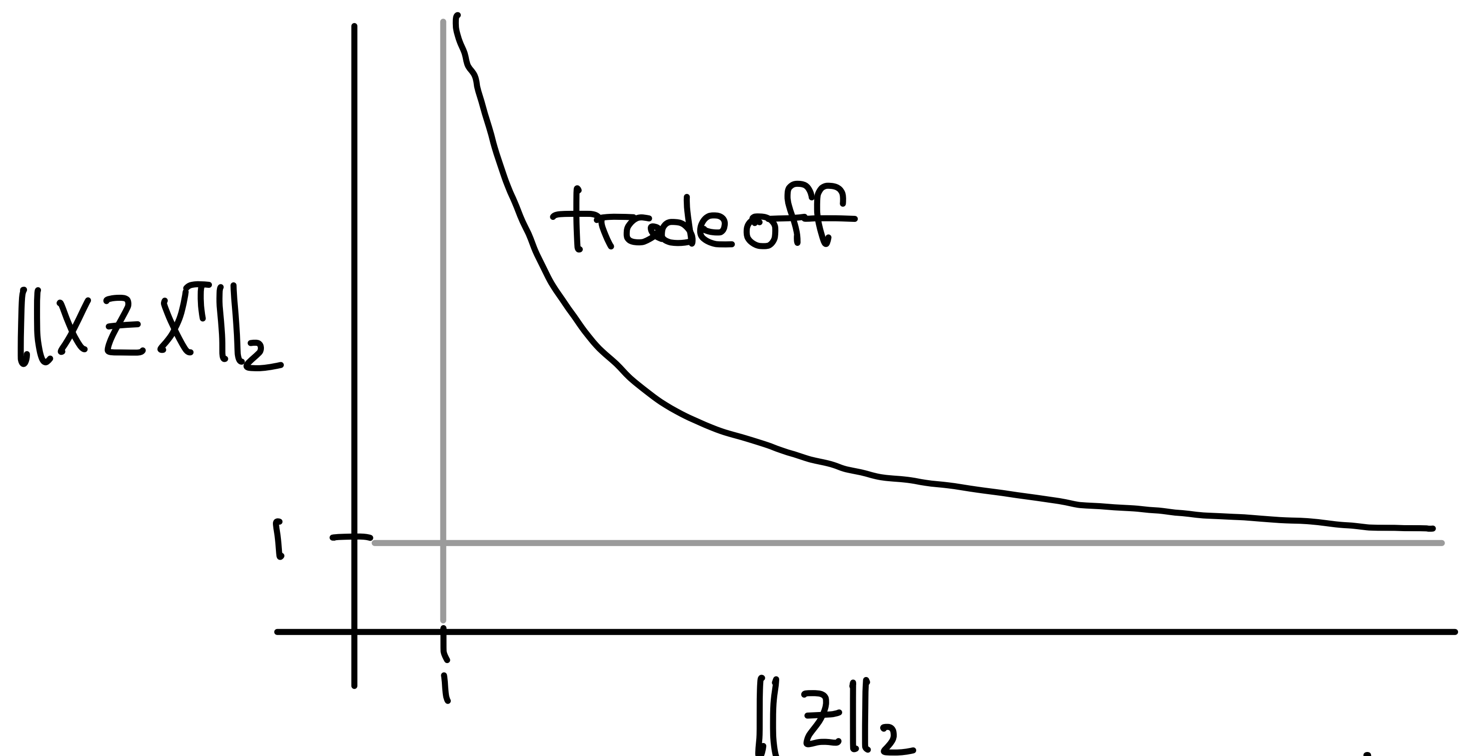
For $x_1, \dots, x_n$ uniform in unit ball in $\mathbb{R}^d$ and best possible matching

$$\mathbb{E}\, \|XZX^T\|_2 \geq \text{const} \cdot \frac{n^{1-2/d}}{d}$$

# Gram-Schmidt Walk Design

For all $\phi \in (0,1]$, obtains

$$\|Z\|_2 \leq \frac{1}{\phi} \qquad \|XZX^T\|_2 \leq \frac{1}{1-\phi}$$




assuming $\|x_i\|_2 \leq 1$

# Gram-Schmidt Walk Design

For all $\phi \in (0,1)$, obtains

$$\|Z\|_2 \leq \frac{1}{\phi} \qquad \|XZX^T\|_2 \leq \frac{1}{1-\phi}$$

$$Z \precsim \left(\phi I + (1-\phi) X^T X\right)^{-1}$$

$$\mathrm{MSE}(\hat{\tau}) \leq \frac{1}{n^2} \min_{\beta \in \mathbb{R}^d} \left[\frac{1}{\phi}\|\mu - X_\beta^T\|^2 + \frac{1}{1-\phi}\|\beta\|^2\right]$$

is a ridge regression
regularized least squares

# Gram-Schmidt Walk Design

small when
$\mu$ near span
of $X$

$$\mathrm{MSE}(\hat{\tau}) \leq \frac{1}{n^2} \min_{\beta \in \mathbb{R}^d} \left[ \frac{1}{\phi} \| \mu - X^T \beta \|^2 + \frac{1}{1-\phi} \| \beta \|^2 \right]$$

Grows
with $n$

Fixed as
$n$ grows

In asymptotic regimes
for $n \gg d$, want $\phi \to 1$

Almost as robust as iid,
but can have much lower MSE

# Gram-Schmidt Walk Design

For all $\phi \in (0,1]$, obtains

$$\| Z \|_2 \leq \frac{1}{\phi} \qquad \| X Z X^T \|_2 \leq \frac{1}{1-\phi}$$

$$\mathrm{MSE}(\hat{\tau}) \leq \frac{1}{n^2} \min_{\beta \in \mathbb{R}^d} \left[ \frac{1}{\phi} \| \mu - X^T \beta \|^2 + \frac{1}{1-\phi} \| \beta \|^2 \right]$$

With sub-Gaussian tails

# Gram-Schmidt Walk of
## Bansal, Dadush, Garg, Lovett '19

A polynomial-time algorithm
that on input $x_1, \dots, x_n$ with $\|x_i\|_2 \le 1$
chooses a random $z \in \{\pm 1\}^n$

so that $\| X Z X^T \|_2 \le 40$

$X_z = \sum_i z(i) x_i$ is $\sqrt{40}$ - subgaussian

→ Polynomial-time algorithm for
Banaszczyk's Discrepancy Theorem

# Gram-Schmidt Walk of
## Bansal, Dadush, Garg, Lovett '19

A polynomial-time algorithm that on input $x_1, \dots, x_n$ with $\|x_i\|_2 \leq 1$ chooses a random $z \in \{\pm 1\}^n$

so that $\| X Z X^T \|_2 \leq \cancel{40}$ $1$

$$X_z = \sum_i z(i) x_i \text{ is } \cancel{\sqrt{40}}\ 1\text{-subgaussian}$$

Harshaw, Sävje, S, Zhang

# Gram-Schmidt Walk Design

Choose tradeoff parameter $\phi \in (0,1)$

Run GSW on augmented covariate vectors

$$b_i = \begin{pmatrix} \sqrt{\phi}\, e_i \\ \sqrt{1-\phi}\, x_i \end{pmatrix}$$

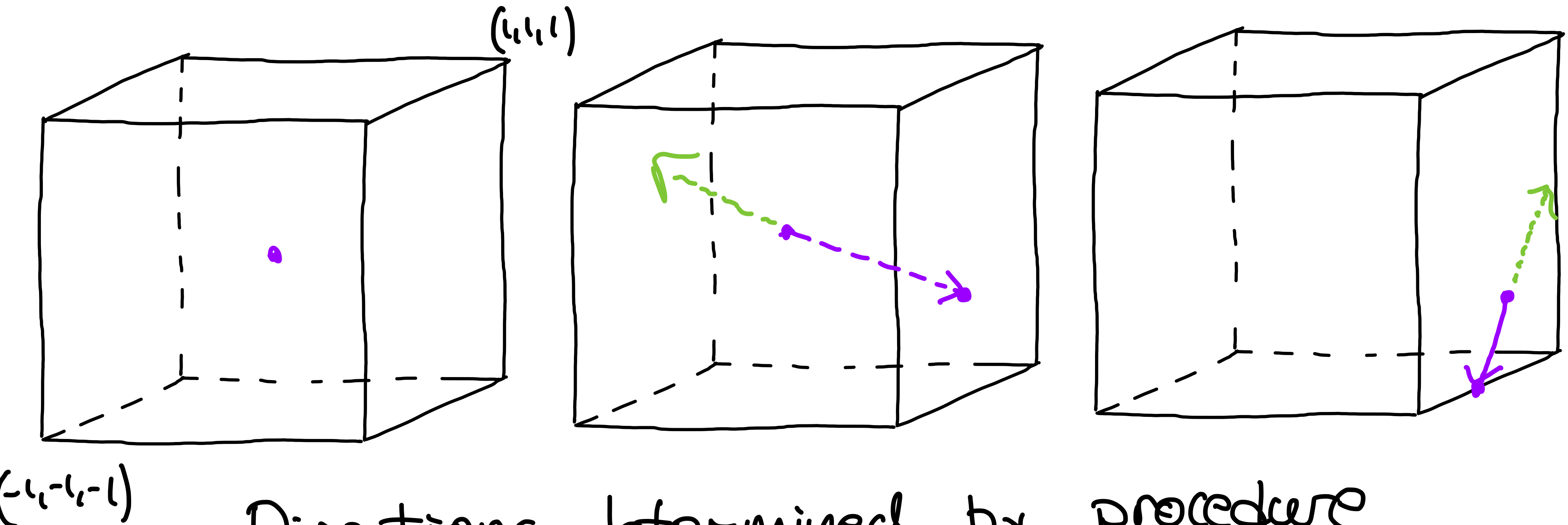
elementary unit vector

$$\text{Input matrix} = \begin{pmatrix} \sqrt{\phi}\, I \\ \sqrt{1-\phi}\, X \end{pmatrix}$$

And, randomly permute their order.*

# Gram-Schmidt Walk Algorithm (BDGL)

Start with $z = (0, \ldots, 0)$. Move inside $[-1,1]^n$




Directions determined by procedure like Gram-Schmidt.

Random steps provide Martingale

See also Beck-Fiala '81, Deville-Tillé '04

# Improving GSW would require assumptions

Tradeoff within constant of optimal

Even for $d$ fixed, $n \to \infty$, $x_i$ from nice distribution

$$\|XZX^T\|_2 \gtrsim 1$$

Zhang '22: $\exists c > 0$ s.t. NP-hard to distinguish

$$\exists Z \quad \|XZX^T\|_2 = 0$$

from $\forall Z \quad \|XZX^T\|_2 \geq c$

# In Practice

Can choose design after examining $X$
and extensive simulation.

Can include functions of covariates.

Can weight covariates

Use ML, or regression, to adjust $\hat{\tau}$?

# In Practice

Can choose design after examining $X$
and extensive simulation.

Can include functions of covariates.

Can weight covariates

Use ML, or regression, to adjust $\hat{\tau}$?

$$\mathrm{MSE}(\hat{\tau}) \leq \frac{1}{n^2} \min_{\beta \in \mathbb{R}^d} \left[ \frac{1}{\phi} \| \mu - X^T \beta \|^2 + \frac{1}{1-\phi} \| \beta \|^2 \right]$$

# Open Problems

Online version (see Kulkarni-Reis-Rothvoss)

Other relations between $X$ and $\mu$?

Matrix Spencer Conjecture

Komlos Conjecture: $\exists C$ s.t.

$\forall x_1, \ldots, x_n$ with $\|x_i\|_2 \leq 1$ for all $i$

$\exists z \in \{\pm 1\}^n$ s.t. $\|\sum z(i) x_i\|_\infty \leq C$

# To learn more...

Cornell University

arXiv > stat > arXiv:1911.03071

Search

Help |

Statistics > Methodology

*[Submitted on 8 Nov 2019 (v1), last revised 4 Feb 2022 (this version, v5)]*

**Balancing covariates in randomized experiments with the Gram–Schmidt Walk design**

Christopher Harshaw, Fredrik Sävje, Daniel Spielman, Peng Zhang

v1: CS style exposition

v2: For experimentalists and statisticians
more results
different simulations

v3, v4, v5, v6: adding stuff referees request

v7: CLT, variance estimator, CI, +

Packages for Julia and R on GitHub