A Unified Framework for Rerandomization using Quadratic Forms

Design and Analysis of Experiments (2024)

Kyle Schindl[1] Zach Branson[2]

Department of Statistics & Data Science, Carnegie Mellon University

[1]schindl@stat.cmu.edu

[2]zach@stat.cmu.edu

Introduction

Problem Setting

Problem Setting

- We focus on treatment vs. control randomized experiments

Problem Setting

- We focus on treatment vs. control randomized experiments
- We are interested in estimating the Average Treatment Effect (ATE)

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

where $Y(1)$ and $Y(0)$ are the outcomes under treatment and control.

Problem Setting

- We focus on treatment vs. control randomized experiments
- We are interested in estimating the Average Treatment Effect (ATE)

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

  where $Y(1)$ and $Y(0)$ are the outcomes under treatment and control.
- In completely randomized experiments, estimates of the ATE are unbiased.

Problem Setting

- We focus on treatment vs. control randomized experiments
- We are interested in estimating the Average Treatment Effect (ATE)

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

  where $Y(1)$ and $Y(0)$ are the outcomes under treatment and control.
- In completely randomized experiments, estimates of the ATE are unbiased.
- Covariate distributions are "balanced" *on average*.

Problem Setting

- We focus on treatment vs. control randomized experiments
- We are interested in estimating the Average Treatment Effect (ATE)

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

  where $Y(1)$ and $Y(0)$ are the outcomes under treatment and control.
- In completely randomized experiments, estimates of the ATE are unbiased.
- Covariate distributions are "balanced" *on average*.
- **We can do better than average.**

Problem Setting

- We focus on treatment vs. control randomized experiments
- We are interested in estimating the Average Treatment Effect (ATE)

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

  where $Y(1)$ and $Y(0)$ are the outcomes under treatment and control.
- In completely randomized experiments, estimates of the ATE are unbiased.
- Covariate distributions are "balanced" *on average*.
- **We can do better than average.**
- By ensuring covariate distributions are closer, we can obtain more precise estimates of the ATE while maintaining unbiasedness.

The Problem of Covariate Imbalance

The Problem of Covariate Imbalance

- Covariate imbalances between treatment and control increases standard errors and reduces statistical power when estimating causal effects (Senn 1989; Lachin 1988; Lin, Zhu, and Su 2015; Branson, Li, and Ding 2023).

The Problem of Covariate Imbalance

- Covariate imbalances between treatment and control increases standard errors and reduces statistical power when estimating causal effects (Senn 1989; Lachin 1988; Lin, Zhu, and Su 2015; Branson, Li, and Ding 2023).
- This is why classical experimental designs suggest blocking similar subjects together when randomizing.

The Problem of Covariate Imbalance

- Covariate imbalances between treatment and control increases standard errors and reduces statistical power when estimating causal effects (Senn 1989; Lachin 1988; Lin, Zhu, and Su 2015; Branson, Li, and Ding 2023).
- This is why classical experimental designs suggest blocking similar subjects together when randomizing.
- Another method of reducing covariate imbalance is to rerandomize.

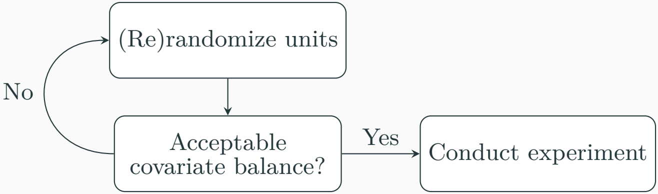


Figure 1: Procedure for implementing rerandomization

Review of Current Rerandomization Methods

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

  $$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where

  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where
  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.
  - $\Sigma := \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where

  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.
  - $\Sigma := \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$
- There have been many extensions to other design settings such as

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where
  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.
  - $\Sigma := \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$
- There have been many extensions to other design settings such as
  - 2^K factorial designs (Branson, Dasgupta, and Rubin 2016; Li, Ding, and Rubin 2020)

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where
  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.
  - $\Sigma := \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$
- There have been many extensions to other design settings such as
  - 2^K factorial designs (Branson, Dasgupta, and Rubin 2016; Li, Ding, and Rubin 2020)
  - Sequential designs (Zhou et al. 2018)

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where
  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.
  - $\Sigma := \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$
- There have been many extensions to other design settings such as
  - 2^K factorial designs (Branson, Dasgupta, and Rubin 2016; Li, Ding, and Rubin 2020)
  - Sequential designs (Zhou et al. 2018)
  - Clustered experiments (Lu et al. 2023)

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where

  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.
  - $\Sigma := \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$
- There have been many extensions to other design settings such as
  - 2^K factorial designs (Branson, Dasgupta, and Rubin 2016; Li, Ding, and Rubin 2020)
  - Sequential designs (Zhou et al. 2018)
  - Clustered experiments (Lu et al. 2023)
  - Bayesian designs (Liu et al. 2023)

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where
  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.
  - $\Sigma := \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$
- There have been many extensions to other design settings such as
  - 2^K factorial designs (Branson, Dasgupta, and Rubin 2016; Li, Ding, and Rubin 2020)
  - Sequential designs (Zhou et al. 2018)
  - Clustered experiments (Lu et al. 2023)
  - Bayesian designs (Liu et al. 2023)
- Most of these extensions utilize the Mahalanobis distance.

Review of Current Rerandomization Methods

- (Morgan and Rubin 2012): Randomize until $M < a$, for some pre-specified $a > 0$, where

$$M := (\overline{X}_T - \overline{X}_C)^T \Sigma^{-1} (\overline{X}_T - \overline{X}_C).$$

  where
  - $\overline{X}_T, \overline{X}_C \in \mathbb{R}^d$ are the covariate means in treatment and control.
  - $\Sigma := \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$
- There have been many extensions to other design settings such as
  - 2^K factorial designs (Branson, Dasgupta, and Rubin 2016; Li, Ding, and Rubin 2020)
  - Sequential designs (Zhou et al. 2018)
  - Clustered experiments (Lu et al. 2023)
  - Bayesian designs (Liu et al. 2023)
- Most of these extensions utilize the Mahalanobis distance.
- We'll consider: What distances are optimal for rerandomization?

Properties of Mahalanobis Rerandomization

- As long as $\sum_{i=1}^{n} W_i = \sum_{i=1}^{n}(1 - W_i)$, where $W_i \in \{0, 1\}$ is the treatment assignment, then $\hat{\tau}$ remains unbiased.

Properties of Mahalanobis Rerandomization

- As long as $\sum_{i=1}^{n} W_i = \sum_{i=1}^{n}(1 - W_i)$, where $W_i \in \{0, 1\}$ is the treatment assignment, then $\hat{\tau}$ remains unbiased.
- If $\overline{X}_T - \overline{X}_C \mid x \sim \mathcal{N}(0, \Sigma)$ then, Mahalanobis Rerandomization applies an equal-percentage variance reduction to all covariates,

$$\mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x, M \leq a) = v_a \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$$

where $v_a = {}^{\mathbb{P}(\chi^2_{d+2} \leq a)}/_{\mathbb{P}(\chi^2_d \leq a)} \leq 1$.

Properties of Mahalanobis Rerandomization

- As long as $\sum_{i=1}^{n} W_i = \sum_{i=1}^{n}(1 - W_i)$, where $W_i \in \{0, 1\}$ is the treatment assignment, then $\hat{\tau}$ remains unbiased.
- If $\overline{X}_T - \overline{X}_C \mid x \sim \mathcal{N}(0, \Sigma)$ then, Mahalanobis Rerandomization applies an equal-percentage variance reduction to all covariates,

$$\mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x, M \leq a) = v_a \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$$

  where $v_a = {}^{\mathbb{P}(\chi^2_{d+2} \leq a)}/_{\mathbb{P}(\chi^2_d \leq a)} \leq 1$.
- When the treatment effect is additive, this leads to a variance reduction of

$$\frac{\mathbb{V}(\hat{\tau} \mid x, M < a)}{\mathbb{V}(\hat{\tau} \mid x)} = 1 - (1 - v_a)R^2 \leq 1$$

  where

$$\hat{\tau} = \frac{1}{n_1} \sum_{i=1}^{n} Y_i(1)\, W_i - \frac{1}{n_0} \sum_{i=1}^{n} Y_i(0)(1 - W_i)$$

Properties of Mahalanobis Rerandomization

- As long as $\sum_{i=1}^n W_i = \sum_{i=1}^n (1 - W_i)$, where $W_i \in \{0,1\}$ is the treatment assignment, then $\hat{\tau}$ remains unbiased.
- If $\overline{X}_T - \overline{X}_C \mid x \sim \mathcal{N}(0, \Sigma)$ then, Mahalanobis Rerandomization applies an equal-percentage variance reduction to all covariates,

$$\mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x, M \leq a) = v_a \mathrm{Cov}(\overline{X}_T - \overline{X}_C \mid x)$$

  where $v_a = {}^{\mathbb{P}(\chi^2_{d+2} \leq a)}/_{\mathbb{P}(\chi^2_d \leq a)} \leq 1$.
- When the treatment effect is additive, this leads to a variance reduction of

$$\frac{\mathbb{V}(\hat{\tau} \mid x, M < a)}{\mathbb{V}(\hat{\tau} \mid x)} = 1 - (1 - v_a) R^2 \leq 1$$

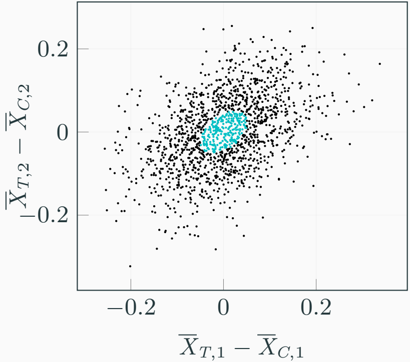
  where

$$\hat{\tau} = \frac{1}{n_1} \sum_{i=1}^n Y_i(1) \, W_i - \frac{1}{n_0} \sum_{i=1}^n Y_i(0)(1 - W_i)$$

- Asymptotically, this is true with unequal sample sizes and non-additivity (Li, Ding, and Rubin 2018).

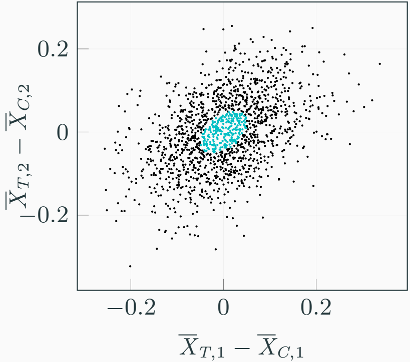
Mahalanobis Rerandomization Visualized

- Geometrically, we can see that Mahalanobis Rerandomization preserves the "shape" of the covariance matrix but reduces the scale.



Mahalanobis Rerandomization Visualized

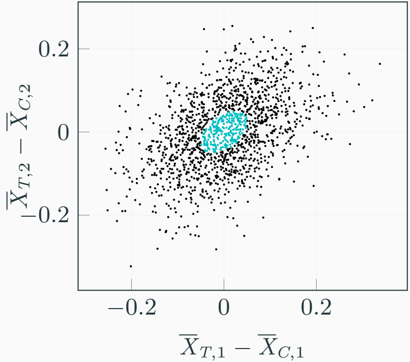
- Geometrically, we can see that Mahalanobis Rerandomization preserves the "shape" of the covariance matrix but reduces the scale.



- Each eigenvector of the covariance matrix Σ is scaled by v_a.

Mahalanobis Rerandomization Visualized

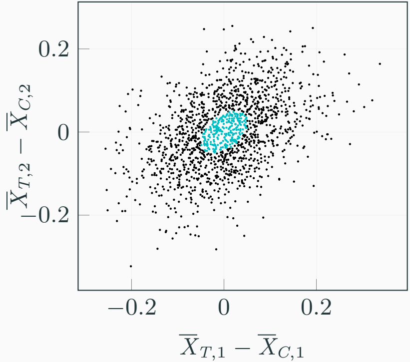
- Geometrically, we can see that Mahalanobis Rerandomization preserves the "shape" of the covariance matrix but reduces the scale.



- Each eigenvector of the covariance matrix Σ is scaled by v_a.
- Placing equal priority on all covariates create challenges since $v_a \to 1$ as $d \to \infty$ (no variance reduction).

Mahalanobis Rerandomization Visualized

- Geometrically, we can see that Mahalanobis Rerandomization preserves the "shape" of the covariance matrix but reduces the scale.



- Each eigenvector of the covariance matrix Σ is scaled by v_a.
- Placing equal priority on all covariates create challenges since $v_a \to 1$ as $d \to \infty$ (no variance reduction).
- Mahalanobis Rerandomization may be "wasting" variance reduction on unimportant covariates

Balancing Metrics as Quadratic Forms

- Other recent proposed balance metrics include:

Balancing Metrics as Quadratic Forms

- Other recent proposed balance metrics include:
- A penalized Mahalanobis distance (Branson and Shao 2021):

$$M_\lambda = (\overline{X}_T - \overline{X}_C)^T (\Sigma + \lambda I_d)^{-1} (\overline{X}_T - \overline{X}_C)$$

Balancing Metrics as Quadratic Forms

- Other recent proposed balance metrics include:
- A penalized Mahalanobis distance (Branson and Shao 2021):

$$M_\lambda = (\overline{X}_T - \overline{X}_C)^T (\Sigma + \lambda I_d)^{-1} (\overline{X}_T - \overline{X}_C)$$

- The Mahalanobis distance after dropping the bottom $d - k$ principal components (Zhang, Yin, and Rubin 2023):

$$M_k = (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)})^T \Sigma_Z^{-1} (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)})$$

where Z are the principal components.

Balancing Metrics as Quadratic Forms

- Other recent proposed balance metrics include:
- A penalized Mahalanobis distance (Branson and Shao 2021):

$$M_\lambda = (\overline{X}_T - \overline{X}_C)^T(\Sigma + \lambda I_d)^{-1}(\overline{X}_T - \overline{X}_C)$$

- The Mahalanobis distance after dropping the bottom $d-k$ principal components (Zhang, Yin, and Rubin 2023):

$$M_k = (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)})^T\Sigma_Z^{-1}(\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)})$$

  where Z are the principal components.
- Each method can be written as $(\overline{X}_T - \overline{X}_C)^T A(\overline{X}_T - \overline{X}_C)$ for some $A \in \mathbb{R}^{d\times d}$, e.g.,

$$M = (\overline{X}_T - \overline{X}_C)^T\underbrace{\Sigma^{-1}}_{A}(\overline{X}_T - \overline{X}_C)$$

Balancing Metrics as Quadratic Forms

- Other recent proposed balance metrics include:
- A penalized Mahalanobis distance (Branson and Shao 2021):

$$M_\lambda = (\overline{X}_T - \overline{X}_C)^T \underbrace{(\Sigma + \lambda I_d)^{-1}}_{A} (\overline{X}_T - \overline{X}_C)$$

- The Mahalanobis distance after dropping the bottom $d - k$ principal components (Zhang, Yin, and Rubin 2023):

$$M_k = (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)})^T \Sigma_Z^{-1} (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)})$$

  where Z are the principal components.
- Each method can be written as $(\overline{X}_T - \overline{X}_C)^T A (\overline{X}_T - \overline{X}_C)$ for some $A \in \mathbb{R}^{d \times d}$, e.g.,

$$M = (\overline{X}_T - \overline{X}_C)^T \underbrace{\Sigma^{-1}}_{A} (\overline{X}_T - \overline{X}_C)$$

Balancing Metrics as Quadratic Forms

- Other recent proposed balance metrics include:
- A penalized Mahalanobis distance (Branson and Shao 2021):

$$M_\lambda = (\overline{X}_T - \overline{X}_C)^T \underbrace{(\Sigma + \lambda I_d)^{-1}}_{A} (\overline{X}_T - \overline{X}_C)$$

- The Mahalanobis distance after dropping the bottom $d - k$ principal components (Zhang, Yin, and Rubin 2023):

$$\begin{aligned} M_k &= (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)})^T \Sigma_Z^{-1} (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)}) \\ &= (\overline{X}_T - \overline{X}_C)^T \underbrace{V \begin{pmatrix} D_k & 0 \\ 0 & 0 \end{pmatrix} V^T}_{A} (\overline{X}_T - \overline{X}_C) \end{aligned}$$

  where D_k are the singular values from $x = UDV^T$.
- Each method can be written as $(\overline{X}_T - \overline{X}_C)^T A (\overline{X}_T - \overline{X}_C)$ for some $A \in \mathbb{R}^{d \times d}$, e.g.,

$$M = (\overline{X}_T - \overline{X}_C)^T \underbrace{\Sigma^{-1}}_{A} (\overline{X}_T - \overline{X}_C)$$

Balancing Metrics as Quadratic Forms

- Other recent proposed balance metrics include:
- A penalized Mahalanobis distance (Branson and Shao 2021):

$$M_\lambda = (\overline{X}_T - \overline{X}_C)^T \underbrace{(\Sigma + \lambda I_d)^{-1}}_{A} (\overline{X}_T - \overline{X}_C)$$

- The Mahalanobis distance after dropping the bottom $d - k$ principal components (Zhang, Yin, and Rubin 2023):

$$\begin{aligned} M_k &= (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)})^T \Sigma_Z^{-1} (\overline{Z}_T^{(k)} - \overline{Z}_C^{(k)}) \\ &= (\overline{X}_T - \overline{X}_C)^T \underbrace{V \begin{pmatrix} D_k & 0 \\ 0 & 0 \end{pmatrix} V^T}_{A} (\overline{X}_T - \overline{X}_C) \end{aligned}$$

  where D_k are the singular values from $x = UDV^T$.
- Each method can be written as $(\overline{X}_T - \overline{X}_C)^T A (\overline{X}_T - \overline{X}_C)$ for some $A \in \mathbb{R}^{d \times d}$, e.g.,

$$M = (\overline{X}_T - \overline{X}_C)^T \underbrace{\Sigma^{-1}}_{A} (\overline{X}_T - \overline{X}_C)$$

- This leads to the question: what is a good choice of A?

Rerandomization using Quadratic Forms

Rerandomization using Quadratic Forms

- In this work, we establish a general framework for randomizing until $Q_A(x) \leq a$ where

$$Q_A(x) := (\overline{X}_T - \overline{X}_C)^T A (\overline{X}_T - \overline{X}_C)$$

Rerandomization using Quadratic Forms

- In this work, we establish a general framework for randomizing until $Q_A(x) \leq a$ where

$$Q_A(x) := (\overline{X}_T - \overline{X}_C)^T A(\overline{X}_T - \overline{X}_C)$$

- We select a by setting an acceptance probability α such that

$$\mathbb{P}(Q_A(x) \leq a) = \alpha$$

Rerandomization using Quadratic Forms

- In this work, we establish a general framework for randomizing until $Q_A(x) \leq a$ where

$$Q_A(x) := (\overline{X}_T - \overline{X}_C)^T A(\overline{X}_T - \overline{X}_C)$$

- We select a by setting an acceptance probability α such that

$$\mathbb{P}(Q_A(x) \leq a) = \alpha$$

- Some key results we cover:

Rerandomization using Quadratic Forms

- In this work, we establish a general framework for randomizing until $Q_A(x) \leq a$ where

$$Q_A(x) := (\overline{X}_T - \overline{X}_C)^T A(\overline{X}_T - \overline{X}_C)$$

- We select a by setting an acceptance probability α such that

$$\mathbb{P}(Q_A(x) \leq a) = \alpha$$

- Some key results we cover:
  - The covariance of $\overline{X}_T - \overline{X}_C$ conditioning on any quadratic form.

Rerandomization using Quadratic Forms

- In this work, we establish a general framework for randomizing until $Q_A(x) \leq a$ where

$$Q_A(x) := (\overline{X}_T - \overline{X}_C)^T A(\overline{X}_T - \overline{X}_C)$$

- We select a by setting an acceptance probability α such that

$$\mathbb{P}(Q_A(x) \leq a) = \alpha$$

- Some key results we cover:
  - The covariance of $\overline{X}_T - \overline{X}_C$ conditioning on any quadratic form.
  - The optimal choice of A for covariate balance

Rerandomization using Quadratic Forms

- In this work, we establish a general framework for randomizing until $Q_A(x) \leq a$ where

$$Q_A(x) := (\overline{X}_T - \overline{X}_C)^T A(\overline{X}_T - \overline{X}_C)$$

- We select a by setting an acceptance probability α such that

$$\mathbb{P}(Q_A(x) \leq a) = \alpha$$

- Some key results we cover:
  - The covariance of $\overline{X}_T - \overline{X}_C$ conditioning on any quadratic form.
  - The optimal choice of A for covariate balance
  - The optimal choice of A for reducing the variance of $\hat{\tau}$

Rerandomization using Quadratic Forms

- In this work, we establish a general framework for randomizing until $Q_A(x) \leq a$ where

$$Q_A(x) := (\overline{X}_T - \overline{X}_C)^T A (\overline{X}_T - \overline{X}_C)$$

- We select a by setting an acceptance probability α such that

$$\mathbb{P}(Q_A(x) \leq a) = \alpha$$

- Some key results we cover:
  - The covariance of $\overline{X}_T - \overline{X}_C$ conditioning on any quadratic form.
  - The optimal choice of A for covariate balance
  - The optimal choice of A for reducing the variance of $\hat{\tau}$
- Throughout, we assume A is positive-definite and that $\overline{X}_T - \overline{X}_C \mid x \sim \mathcal{N}(0, \Sigma)$.

Rerandomization using Quadratic Forms

- In this work, we establish a general framework for randomizing until $Q_A(x) \leq a$ where

$$Q_A(x) := (\overline{X}_T - \overline{X}_C)^T A(\overline{X}_T - \overline{X}_C)$$

- We select a by setting an acceptance probability α such that

$$\mathbb{P}(Q_A(x) \leq a) = \alpha$$

- Some key results we cover:
  - The covariance of $\overline{X}_T - \overline{X}_C$ conditioning on any quadratic form.
  - The optimal choice of A for covariate balance
  - The optimal choice of A for reducing the variance of $\hat{\tau}$
- Throughout, we assume A is positive-definite and that $\overline{X}_T - \overline{X}_C \mid x \sim \mathcal{N}(0, \Sigma)$.
- Also note that we assume the covariates have been standardized.

Covariance Reduction under Quadratic Form Rerandomization

Theorem 1 (Covariance Reduction)
Suppose that $\Sigma^{1/2}A\Sigma^{1/2}$ and Σ share an eigenbasis. Then,

$$Cov\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right) = \Sigma^{1/2}\Gamma\Big(diag\{(q_{j,\eta})_{1\leq j\leq d}\}\Big)\Gamma^T\Sigma^{1/2}$$

Covariance Reduction under Quadratic Form Rerandomization

Theorem 1 (Covariance Reduction)
Suppose that $\Sigma^{1/2} A \Sigma^{1/2}$ and Σ share an eigenbasis. Then,

$$Cov\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right) = \Sigma^{1/2}\Gamma\Big(diag\{(q_{j,\eta})_{1 \leq j \leq d}\}\Big)\Gamma^T\Sigma^{1/2}$$

- $\Gamma \in \mathbb{R}^{d \times d}$ is the orthogonal matrix of eigenvectors of Σ

Covariance Reduction under Quadratic Form Rerandomization

Theorem 1 (Covariance Reduction)
Suppose that $\Sigma^{1/2} A \Sigma^{1/2}$ *and* Σ *share an eigenbasis. Then,*

$$Cov\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \le a\right) = \Sigma^{1/2}\Gamma\Big(diag\{(q_{j,\eta})_{1\le j\le d}\}\Big)\Gamma^T\Sigma^{1/2}$$

- $\Gamma \in \mathbb{R}^{d\times d}$ is the orthogonal matrix of eigenvectors of Σ
- $q_{j,\eta}$ is defined as

$$q_{j,\eta} = \mathbb{E}\left[\mathcal{Z}_j^2 \mid x, \sum_{j=1}^{d} \eta_j \mathcal{Z}_j^2 \le a\right]$$

Covariance Reduction under Quadratic Form Rerandomization

Theorem 1 (Covariance Reduction)
Suppose that $\Sigma^{1/2} A \Sigma^{1/2}$ and Σ share an eigenbasis. Then,

$$Cov\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right) = \Sigma^{1/2}\Gamma\Big(diag\{(q_{j,\eta})_{1 \leq j \leq d}\}\Big)\Gamma^T\Sigma^{1/2}$$

- $\Gamma \in \mathbb{R}^{d \times d}$ is the orthogonal matrix of eigenvectors of Σ
- $q_{j,\eta}$ is defined as

$$q_{j,\eta} = \mathbb{E}\left[\mathcal{Z}_j^2 \mid x, \sum_{j=1}^{d} \eta_j \mathcal{Z}_j^2 \leq a\right]$$

- $\eta_1 \geq \cdots \geq \eta_d$ are the eigenvalues of $\Sigma^{1/2} A \Sigma^{1/2}$

Covariance Reduction under Quadratic Form Rerandomization

Theorem 1 (Covariance Reduction)
Suppose that $\Sigma^{1/2} A \Sigma^{1/2}$ and Σ share an eigenbasis. Then,

$$Cov\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right) = \Sigma^{1/2}\Gamma\Big(diag\{(q_{j,\eta})_{1 \leq j \leq d}\}\Big)\Gamma^T\Sigma^{1/2}$$

- $\Gamma \in \mathbb{R}^{d \times d}$ is the orthogonal matrix of eigenvectors of Σ
- $q_{j,\eta}$ is defined as

$$q_{j,\eta} = \mathbb{E}\left[\mathcal{Z}_j^2 \mid x, \sum_{j=1}^{d} \eta_j \mathcal{Z}_j^2 \leq a\right]$$

- $\eta_1 \geq \cdots \geq \eta_d$ are the eigenvalues of $\Sigma^{1/2} A \Sigma^{1/2}$
- $\mathcal{Z}_1, \ldots, \mathcal{Z}_d \sim \mathcal{N}(0, 1)$

Covariance Reduction under Quadratic Form Rerandomization

Theorem 1 (Covariance Reduction)
Suppose that $\Sigma^{1/2} A \Sigma^{1/2}$ and Σ share an eigenbasis. Then,

$$Cov\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right) = \Sigma^{1/2}\Gamma\Big(diag\{(q_{j,\eta})_{1\leq j\leq d}\}\Big)\Gamma^T\Sigma^{1/2}$$

- $\Gamma \in \mathbb{R}^{d\times d}$ is the orthogonal matrix of eigenvectors of Σ
- $q_{j,\eta}$ is defined as

$$q_{j,\eta} = \mathbb{E}\left[\mathcal{Z}_j^2 \mid x, \sum_{j=1}^{d} \eta_j \mathcal{Z}_j^2 \leq a\right]$$

- $\eta_1 \geq \cdots \geq \eta_d$ are the eigenvalues of $\Sigma^{1/2} A \Sigma^{1/2}$
- $\mathcal{Z}_1, \ldots, \mathcal{Z}_d \sim \mathcal{N}(0, 1)$
- $q_{1,\eta} \leq \cdots \leq q_{d,\eta} \leq 1$.

Covariance Reduction under Quadratic Form Rerandomization

Theorem 1 (Covariance Reduction)
Suppose that $\Sigma^{1/2} A \Sigma^{1/2}$ and Σ share an eigenbasis. Then,

$$Cov\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right) = \Sigma^{1/2}\Gamma\Big(diag\{(q_{j,\eta})_{1\leq j\leq d}\}\Big)\Gamma^T\Sigma^{1/2}$$

- $\Gamma \in \mathbb{R}^{d\times d}$ is the orthogonal matrix of eigenvectors of Σ
- $q_{j,\eta}$ is defined as

$$\begin{aligned} q_{j,\eta} &= \mathbb{E}\left[\mathcal{Z}_j^2 \mid x, \sum_{j=1}^{d} \eta_j \mathcal{Z}_j^2 \leq a\right] \\ &= \frac{p_d}{\eta_j}\det(\Sigma^{1/2}A\Sigma^{1/2})^{1/d}\alpha^{2/d} + o(\alpha^{2/d}) \end{aligned}$$

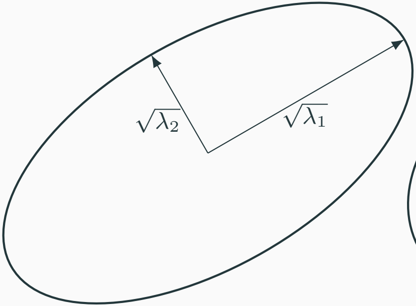
  for a sufficiently small α, where $p_d = \frac{2\pi}{d+2}\left(\frac{2\pi^{d/2}}{d\Gamma(d/2)}\right)^{-2/d}$ (Lu et al. 2023).

- $\eta_1 \geq \cdots \geq \eta_d$ are the eigenvalues of $\Sigma^{1/2}A\Sigma^{1/2}$
- $\mathcal{Z}_1, \ldots, \mathcal{Z}_d \sim \mathcal{N}(0, 1)$
- $q_{1,\eta} \leq \cdots \leq q_{d,\eta} \leq 1$.

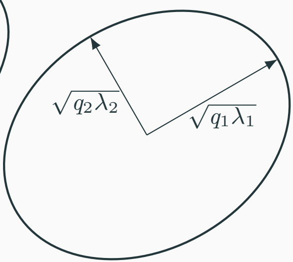
Geometric Intuition behind Quadratic Form Rerandomization

Geometric Intuition behind Quadratic Form Rerandomization

- The eigenvalues of $\mathrm{Cov}\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right)$ are given by $q_{1,\eta}\lambda_1, \ldots, q_{d,\eta}\lambda_d$.



(a) Covariance Matrix before rerandomization

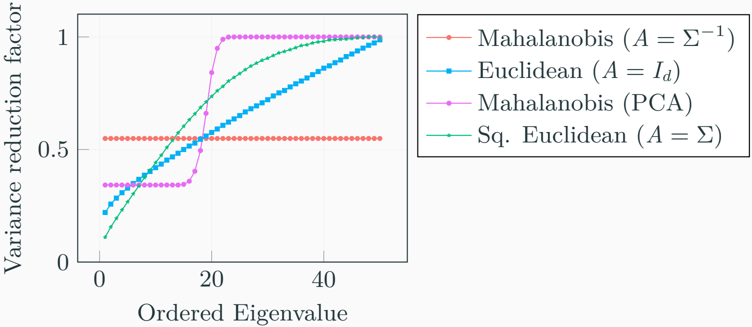


(b) Covariance Matrix after rerandomization

Variance Reduction Factors Visualized

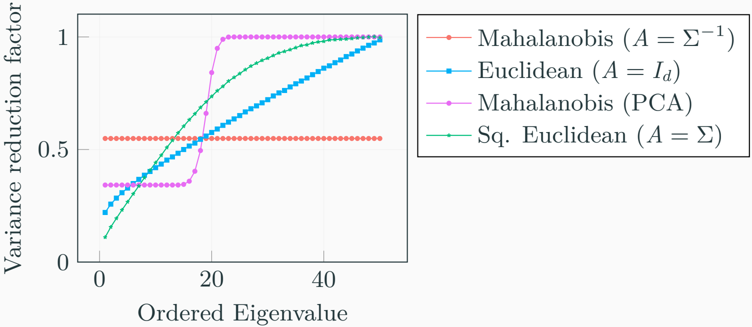
Variance Reduction Factors Visualized

- Depending on the choice of A, there is a different amount of variance reduction applied to each eigenvalue.



Variance Reduction Factors Visualized

- Depending on the choice of A, there is a different amount of variance reduction applied to each eigenvalue.



- We show that different choices of A can lead to significantly different precision levels when estimating $\hat{\tau}$.

Variance Reduction Visualized

Variance Reduction Visualized

- Depending on the relationship between the potential outcomes and x, and the eigenstructure of Σ, different choices of A yield different levels of precision.

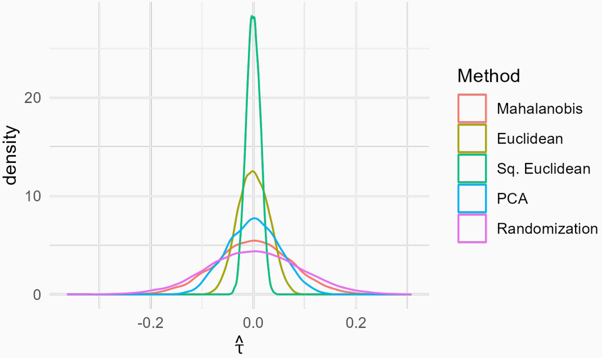


Figure 3: Distribution of $\widehat{\tau}$

Variance Reduction Visualized

- Depending on the relationship between the potential outcomes and x, and the eigenstructure of Σ, different choices of A yield different levels of precision.

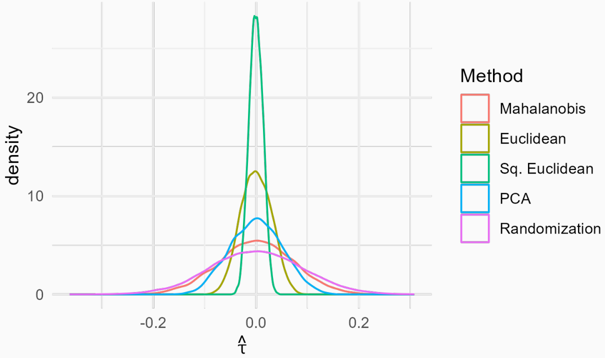


Figure 3: Distribution of $\widehat{\tau}$

- Given the clear differences between each of these methods, the question arises: which choice of A is optimal?

Choosing the Optimal Matrix A

Choosing the Optimal Matrix A

- First we consider optimality in terms of covariate balance.

Choosing the Optimal Matrix A

- First we consider optimality in terms of covariate balance.
- We consider two targets for quantifying covariate balance:

Choosing the Optimal Matrix A

- First we consider optimality in terms of covariate balance.
- We consider two targets for quantifying covariate balance:
    - The choice of A that maximizes the variance reduction applied across all covariates,

$$\max_{A \in \mathbb{R}^{d \times d}} \sum_{j=1}^{d} (1 - q_{j,\eta}) = \min_{A \in \mathbb{R}^{d \times d}} \sum_{j=1}^{d} q_{j,\eta}$$

Choosing the Optimal Matrix A

- First we consider optimality in terms of covariate balance.
- We consider two targets for quantifying covariate balance:
  - The choice of A that maximizes the variance reduction applied across all covariates,

$$\max_{A \in \mathbb{R}^{d \times d}} \sum_{j=1}^{d} (1 - q_{j,\eta}) = \min_{A \in \mathbb{R}^{d \times d}} \sum_{j=1}^{d} q_{j,\eta}$$

  - The choice of A that constrains the covariance matrix after rerandomization the most, which we define as

$$\min_{A \in \mathbb{R}^{d \times d}} ||\mathrm{Cov}\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right)||_F^2$$

  where $|| \cdot ||_F^2 = \sum_{j=1}^{d} \sigma_j(\cdot)$ and $\sigma_j(\cdot)$ are the singular values of the matrix.

Choosing the Optimal Matrix A

- First we consider optimality in terms of covariate balance.
- We consider two targets for quantifying covariate balance:
  - The choice of A that maximizes the variance reduction applied across all covariates,

$$\max_{A \in \mathbb{R}^{d \times d}} \sum_{j=1}^{d} (1 - q_{j,\eta}) = \min_{A \in \mathbb{R}^{d \times d}} \sum_{j=1}^{d} q_{j,\eta}$$

  - The choice of A that constrains the covariance matrix after rerandomization the most, which we define as

$$\min_{A \in \mathbb{R}^{d \times d}} ||\mathrm{Cov}\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right)||_F^2$$

    where $|| \cdot ||_F^2 = \sum_{j=1}^{d} \sigma_j(\cdot)$ and $\sigma_j(\cdot)$ are the singular values of the matrix.
  - Equivalently, $\sigma_j^2(\cdot)$ are the eigenvalues of the sample covariance matrix.

Choosing the Optimal Matrix A

- First we consider optimality in terms of covariate balance.
- We consider two targets for quantifying covariate balance:
  - The choice of A that maximizes the variance reduction applied across all covariates,

$$\max_{A \in \mathbb{R}^{d \times d}} \sum_{j=1}^{d} (1 - q_{j,\eta}) = \min_{A \in \mathbb{R}^{d \times d}} \sum_{j=1}^{d} q_{j,\eta}$$

  - The choice of A that constrains the covariance matrix after rerandomization the most, which we define as

$$\min_{A \in \mathbb{R}^{d \times d}} ||\text{Cov}\left(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a\right)||_F^2$$

    where $|| \cdot ||_F^2 = \sum_{j=1}^{d} \sigma_j(\cdot)$ and $\sigma_j(\cdot)$ are the singular values of the matrix.
  - Equivalently, $\sigma_j^2(\cdot)$ are the eigenvalues of the sample covariance matrix.
- Later, we investigate how this impacts the precision of $\hat{\tau}$.

Optimal Choice of A for covariate balance

Optimal Choice of A for covariate balance

Theorem 2 (Covariate Balance)

For all positive-definite matrices $A \in \mathbb{R}^{d\times d}$

- $A = I_d$ *minimizes the Frobenius norm:*

$$||Cov(\overline{X}_T - \overline{X}_C \mid x, Q_I(x) \leq a)||_F \leq ||Cov(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a)||_F + o(\alpha^{2/d})$$

- $A = \Sigma^{-1}$ *maximizes the total variance reduction applied*

$$\sum_{j=1}^{d} v_a \leq \sum_{j=1}^{d} q_{j,\eta} + o(\alpha^{2/d})$$

Optimal Choice of A for covariate balance

Theorem 2 (Covariate Balance)

For all positive-definite matrices $A \in \mathbb{R}^{d \times d}$

- $A = I_d$ *minimizes the Frobenius norm:*

$$||Cov(\overline{X}_T - \overline{X}_C \mid x, Q_I(x) \leq a)||_F \leq ||Cov(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a)||_F + o(\alpha^{2/d})$$

- $A = \Sigma^{-1}$ *maximizes the total variance reduction applied*

$$\sum_{j=1}^{d} v_a \leq \sum_{j=1}^{d} q_{j,\eta} + o(\alpha^{2/d})$$

- These two methods represent opposite sides of the rerandomization spectrum:

Optimal Choice of A for covariate balance

Theorem 2 (Covariate Balance)
For all positive-definite matrices $A \in \mathbb{R}^{d\times d}$

- $A = I_d$ *minimizes the Frobenius norm:*

$$||Cov(\overline{X}_T - \overline{X}_C \mid x, Q_I(x) \leq a)||_F \leq ||Cov(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a)||_F + o(\alpha^{2/d})$$

- $A = \Sigma^{-1}$ *maximizes the total variance reduction applied*

$$\sum_{j=1}^{d} v_a \leq \sum_{j=1}^{d} q_{j,\eta} + o(\alpha^{2/d})$$

- These two methods represent opposite sides of the rerandomization spectrum:
  - Mahalanobis Rerandomization ($A = \Sigma^{-1}$) leaves the shape of the covariance matrix unchanged by applying an equal variance reduction to all eigenvectors.

Optimal Choice of A for covariate balance

Theorem 2 (Covariate Balance)
For all positive-definite matrices $A \in \mathbb{R}^{d \times d}$

- $A = I_d$ *minimizes the Frobenius norm:*

$$||Cov(\overline{X}_T - \overline{X}_C \mid x, Q_I(x) \leq a)||_F \leq ||Cov(\overline{X}_T - \overline{X}_C \mid x, Q_A(x) \leq a)||_F + o(\alpha^{2/d})$$

- $A = \Sigma^{-1}$ *maximizes the total variance reduction applied*

$$\sum_{j=1}^{d} v_a \leq \sum_{j=1}^{d} q_{j,\eta} + o(\alpha^{2/d})$$

- These two methods represent opposite sides of the rerandomization spectrum:
  - Mahalanobis Rerandomization ($A = \Sigma^{-1}$) leaves the shape of the covariance matrix unchanged by applying an equal variance reduction to all eigenvectors.
  - Euclidean Rerandomization ($A = I_d$) scales each eigenvector by a different factor such that their magnitude is the same.

Quantifying the variance reduction to $\widehat{\tau}$

Quantifying the variance reduction to $\widehat{\tau}$

- When the treatment effect is additive, it follows that

$$Y_i(W_i) = \beta_0 + Z_i\beta_Z + \tau W_i + \varepsilon_i$$

where $\beta_0 + Z_i\beta_Z$ represent the linear projection of Y_i onto the principal components $Z = XV$ and ε_i is a residual that encompasses deviations from the projection.

Quantifying the variance reduction to $\widehat{\tau}$

- When the treatment effect is additive, it follows that

$$Y_i(W_i) = \beta_0 + Z_i\beta_Z + \tau W_i + \varepsilon_i$$

where $\beta_0 + Z_i\beta_Z$ represent the linear projection of Y_i onto the principal components $Z = XV$ and ε_i is a residual that encompasses deviations from the projection.

Theorem 3 (Variance Reduction)
Suppose the treatment effect is additive. Then,

$$\mathbb{V}(\widehat{\tau} \mid x) - \mathbb{V}(\widehat{\tau} \mid x, Q_A(x) \leq a) = \sum_{j=1}^{d} \beta_{Z,j}^2 \lambda_j (1 - q_{j,\eta}) \geq 0$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of Σ and $\hat{\tau}$ is the mean-difference estimator.

The Optimal Choice of A for variance reduction to $\hat{\tau}$

The Optimal Choice of A for variance reduction to $\hat{\tau}$

Theorem 4 (Optimal A)
Suppose that the treatment effect is additive. Then for all positive-definite matrices $A \in \mathbb{R}^{d\times d}$,

$$A^* = \Gamma \begin{pmatrix} \beta_{Z,1}^2 & & 0 \\ & \ddots & \\ 0 & & \beta_{Z,d}^2 \end{pmatrix} \Gamma^T$$

minimizes the variance of the mean-differences estimator, i.e.,

$$\mathbb{V}(\widehat{\tau} \mid x, Q_{A^*}(x) \leq a) \leq \mathbb{V}(\widehat{\tau} \mid x, Q_A(x) \leq a) + o(\alpha^{2/d})$$

where Γ is the orthogonal matrix of eigenvectors of Σ.

Which A is optimal? It depends.

Which A is optimal? It depends.

- For any choice of A there exists some β_Z such that A is optimal.

Which A is optimal? It depends.

- For any choice of A there exists some β_Z such that A is optimal.
- For example:
  - When $\beta_{Z,j}^2 = \frac{1}{\lambda_j}$, $A = \Sigma^{-1}$ (Mahalanobis Rerandomization) is optimal.

Which A is optimal? It depends.

- For any choice of A there exists some β_Z such that A is optimal.
- For example:
  - When $\beta_{Z,j}^2 = \frac{1}{\lambda_j}$, $A = \Sigma^{-1}$ (Mahalanobis Rerandomization) is optimal.
  - When $\beta_{Z,j}^2 = 1$, $A = I_d$ (Euclidean Rerandomization) is optimal.

Which A is optimal? It depends.

- For any choice of A there exists some β_Z such that A is optimal.
- For example:
  - When $\beta_{Z,j}^2 = \frac{1}{\lambda_j}$, $A = \Sigma^{-1}$ (Mahalanobis Rerandomization) is optimal.
  - When $\beta_{Z,j}^2 = 1$, $A = I_d$ (Euclidean Rerandomization) is optimal.
  - When $\beta_{Z,j}^2 = \lambda_j$, the choice $A = \Sigma$ is optimal.

Which A is optimal? It depends.

- For any choice of A there exists some β_Z such that A is optimal.
- For example:
    - When $\beta_{Z,j}^2 = \frac{1}{\lambda_j}$, $A = \Sigma^{-1}$ (Mahalanobis Rerandomization) is optimal.
    - When $\beta_{Z,j}^2 = 1$, $A = I_d$ (Euclidean Rerandomization) is optimal.
    - When $\beta_{Z,j}^2 = \lambda_j$, the choice $A = \Sigma$ is optimal.
- When the top eigenvectors and potential outcomes are closely related, choices like $A = I_d, \Sigma$ or PCA rerandomization tend to be better

Which A is optimal? It depends.

- For any choice of A there exists some β_Z such that A is optimal.
- For example:
  - When $\beta_{Z,j}^2 = \frac{1}{\lambda_j}$, $A = \Sigma^{-1}$ (Mahalanobis Rerandomization) is optimal.
  - When $\beta_{Z,j}^2 = 1$, $A = I_d$ (Euclidean Rerandomization) is optimal.
  - When $\beta_{Z,j}^2 = \lambda_j$, the choice $A = \Sigma$ is optimal.
- When the top eigenvectors and potential outcomes are closely related, choices like $A = I_d, \Sigma$ or PCA rerandomization tend to be better
- When the bottom eigenvectors and potential outcomes are closely related, methods like $A = \Sigma^{-1}$ are better.

Which A is optimal? It depends.

- For any choice of A there exists some β_Z such that A is optimal.
- For example:
  - When $\beta_{Z,j}^2 = \frac{1}{\lambda_j}$, $A = \Sigma^{-1}$ (Mahalanobis Rerandomization) is optimal.
  - When $\beta_{Z,j}^2 = 1$, $A = I_d$ (Euclidean Rerandomization) is optimal.
  - When $\beta_{Z,j}^2 = \lambda_j$, the choice $A = \Sigma$ is optimal.
- When the top eigenvectors and potential outcomes are closely related, choices like $A = I_d, \Sigma$ or PCA rerandomization tend to be better
- When the bottom eigenvectors and potential outcomes are closely related, methods like $A = \Sigma^{-1}$ are better.
- However, different choices may be less risky than others.

The minimax optimal choice of A

Theorem 5 (Minimax A)
The choice $A = I_d$ minimizes the maximum difference between the optimal quadratic form and any other choice of A, given by

$$\min_{A \in \mathbb{R}^{d \times d}} \max_{||\beta_Z||_2 \leq c} \left| \mathbb{V}(\widehat{\tau} \mid x, Q_{A^*}(x) \leq a) - \mathbb{V}(\widehat{\tau} \mid x, Q_A(x) \leq a) \right|.$$

The minimax optimal choice of A

Theorem 5 (Minimax A)
The choice $A = I_d$ minimizes the maximum difference between the optimal quadratic form and any other choice of A, given by

$$\min_{A \in \mathbb{R}^{d\times d}} \max_{||\beta_Z||_2 \le c} \left| \mathbb{V}(\widehat{\tau} \mid x, Q_{A^*}(x) \le a) - \mathbb{V}(\widehat{\tau} \mid x, Q_A(x) \le a) \right| .$$

- Under $A = I_d$, the precision of $\hat{\tau}$ is never too far from the optimal choice.

The minimax optimal choice of A

Theorem 5 (Minimax A)
The choice $A = I_d$ minimizes the maximum difference between the optimal quadratic form and any other choice of A, given by

$$\min_{A \in \mathbb{R}^{d \times d}} \max_{||\beta_Z||_2 \leq c} \left| \mathbb{V}(\widehat{\tau} \mid x, Q_{A^*}(x) \leq a) - \mathbb{V}(\widehat{\tau} \mid x, Q_A(x) \leq a) \right| .$$

- Under $A = I_d$, the precision of $\hat{\tau}$ is never too far from the optimal choice.
- For intuition, compare the variance reduction factors between the two:

$$q_{j,\lambda} = \mathbb{E}\left[\mathcal{Z}_j^2 \mid x, \sum_{j=1}^{d} \lambda_j \mathcal{Z}_j^2 \leq a \right]$$

$$q^*_{j,\beta\lambda} = \mathbb{E}\left[\mathcal{Z}_j^2 \mid x, \sum_{j=1}^{d} \beta^2_{Z,j} \lambda_j \mathcal{Z}_j^2 \leq a \right]$$

Simulations

Simulations

- Because the optimal choice of A depends on β_Z as well as the distribution of eigenvalues, we simulate $\lambda_1, \ldots, \lambda_d \sim d \cdot \text{Dirichlet}(\gamma)$.

Simulations

- Because the optimal choice of A depends on β_Z as well as the distribution of eigenvalues, we simulate $\lambda_1, \ldots, \lambda_d \sim d \cdot \text{Dirichlet}(\gamma)$.
- That way $\sum_{j=1}^{d} \lambda_j = d = \text{tr}(\Sigma)$

Simulations

- Because the optimal choice of A depends on β_Z as well as the distribution of eigenvalues, we simulate $\lambda_1, \ldots, \lambda_d \sim d \cdot \text{Dirichlet}(\gamma)$.
- That way $\sum_{j=1}^d \lambda_j = d = \text{tr}(\Sigma)$
- We simulate $n = 500$ observations where 250 units are sent to treatment and 250 units are sent to control.

Simulations

- Because the optimal choice of A depends on β_Z as well as the distribution of eigenvalues, we simulate $\lambda_1, \ldots, \lambda_d \sim d \cdot \text{Dirichlet}(\gamma)$.
- That way $\sum_{j=1}^{d} \lambda_j = d = \text{tr}(\Sigma)$
- We simulate $n = 500$ observations where 250 units are sent to treatment and 250 units are sent to control.
- Potential outcomes $Y_i(0)$ and $Y_i(1)$ are defined as

$$Y_i(0) = \mathcal{N}(Z\beta_Z, 1)$$
$$Y_i(1) = Y_i(0) + \tau$$

where $\tau = 1$, $\beta_Z \in \mathbb{R}^d$, and $Z = XV$ are the principal components of X.

Simulations

- Because the optimal choice of A depends on β_Z as well as the distribution of eigenvalues, we simulate $\lambda_1, \dots, \lambda_d \sim d \cdot \text{Dirichlet}(\gamma)$.
- That way $\sum_{j=1}^{d} \lambda_j = d = \text{tr}(\Sigma)$
- We simulate $n = 500$ observations where 250 units are sent to treatment and 250 units are sent to control.
- Potential outcomes $Y_i(0)$ and $Y_i(1)$ are defined as

$$Y_i(0) = \mathcal{N}(Z\beta_Z, 1)$$
$$Y_i(1) = Y_i(0) + \tau$$

  where $\tau = 1$, $\beta_Z \in \mathbb{R}^d$, and $Z = XV$ are the principal components of X.
- In this presentation we consider two choices of β_Z:

Simulations

- Because the optimal choice of A depends on β_Z as well as the distribution of eigenvalues, we simulate $\lambda_1, \ldots, \lambda_d \sim d \cdot \text{Dirichlet}(\gamma)$.
- That way $\sum_{j=1}^{d} \lambda_j = d = \text{tr}(\Sigma)$
- We simulate $n = 500$ observations where 250 units are sent to treatment and 250 units are sent to control.
- Potential outcomes $Y_i(0)$ and $Y_i(1)$ are defined as

$$Y_i(0) = \mathcal{N}(Z\beta_Z, 1)$$
$$Y_i(1) = Y_i(0) + \tau$$

  where $\tau = 1$, $\beta_Z \in \mathbb{R}^d$, and $Z = XV$ are the principal components of X.
- In this presentation we consider two choices of β_Z:
  - β_Z such that Mahalanobis is optimal ($A = \Sigma^{-1}$)

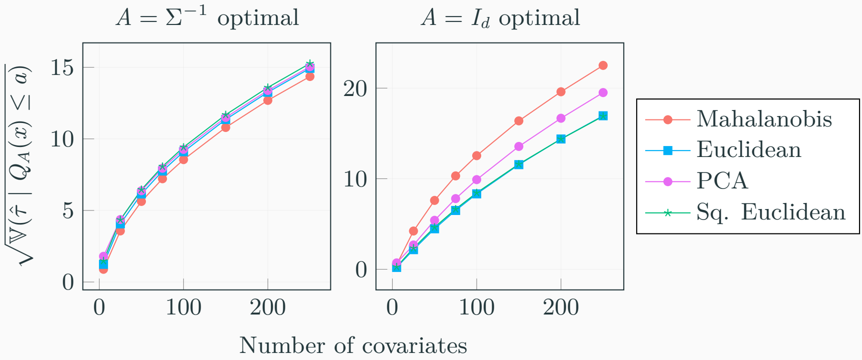
Simulations

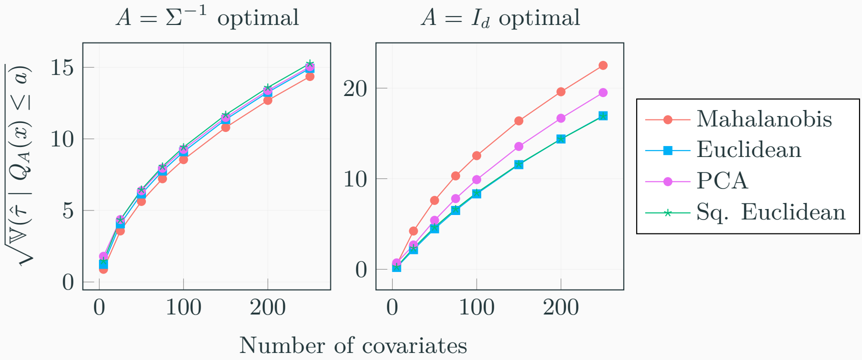
- Because the optimal choice of A depends on β_Z as well as the distribution of eigenvalues, we simulate $\lambda_1, \ldots, \lambda_d \sim d \cdot \text{Dirichlet}(\gamma)$.
- That way $\sum_{j=1}^{d} \lambda_j = d = \text{tr}(\Sigma)$
- We simulate $n = 500$ observations where 250 units are sent to treatment and 250 units are sent to control.
- Potential outcomes $Y_i(0)$ and $Y_i(1)$ are defined as

$$Y_i(0) = \mathcal{N}(Z\beta_Z, 1)$$
$$Y_i(1) = Y_i(0) + \tau$$

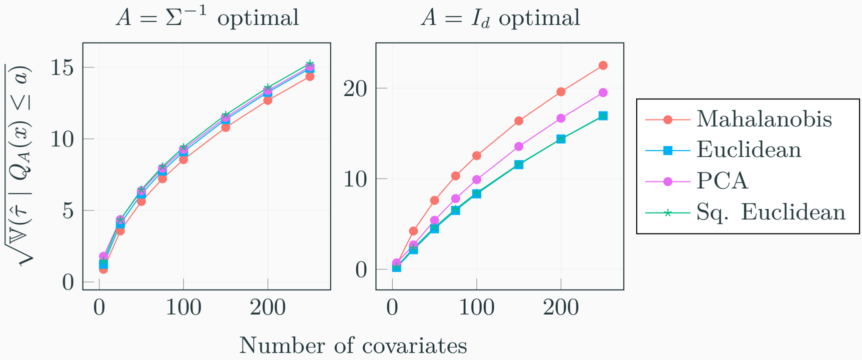
  where $\tau = 1$, $\beta_Z \in \mathbb{R}^d$, and $Z = XV$ are the principal components of X.
- In this presentation we consider two choices of β_Z:
  - β_Z such that Mahalanobis is optimal ($A = \Sigma^{-1}$)
  - β_Z such that Euclidean is optimal ($A = I_d$)

Simulations





- In these settings, Euclidean is either the best or second best choice of A



- In these settings, Euclidean is either the best or second best choice of A
- We tend to find asymmetry: Mahalanobis is either the best in settings favorable to it, or it is the worst.

Conclusion

Conclusion

- We derived general results for rerandomization using any quadratic form.

Conclusion

- We derived general results for rerandomization using any quadratic form.
- This allows researchers to quickly and easily derive new rerandomization schemes not previously considered.

Conclusion

- We derived general results for rerandomization using any quadratic form.
- This allows researchers to quickly and easily derive new rerandomization schemes not previously considered.
- We find that $A = \Sigma^{-1}$ and $A = I_d$ optimize different measurements of covariate balance.

Conclusion

- We derived general results for rerandomization using any quadratic form.
- This allows researchers to quickly and easily derive new rerandomization schemes not previously considered.
- We find that $A = \Sigma^{-1}$ and $A = I_d$ optimize different measurements of covariate balance.
- However, $A = I_d$ is never too far away from the optimal choice, and tends to be more precise on average.

Conclusion

- We derived general results for rerandomization using any quadratic form.
- This allows researchers to quickly and easily derive new rerandomization schemes not previously considered.
- We find that $A = \Sigma^{-1}$ and $A = I_d$ optimize different measurements of covariate balance.
- However, $A = I_d$ is never too far away from the optimal choice, and tends to be more precise on average.
- In future work, we plan to extend this framework to sequential experiments as well as multi-valued treatments.

Thank you!

- Questions?

References

Branson, Zach, Tirthankar Dasgupta, and Donald B Rubin (2016). **"Improving covariate balance in 2 K factorial designs via rerandomization with an application to a New York City Department of Education High School Study".** In.

Branson, Zach, Xinran Li, and Peng Ding (May 2023). **"Power and sample size calculations for rerandomization".** In: *Biometrika* 111.1, pp. 355–363. ISSN: 1464-3510. eprint: *https://academic.oup.com/biomet/article-pdf/111/1/355/56665596/asad027.pdf*.

Branson, Zach and Stephane Shao (2021). **"Ridge rerandomization: An experimental design strategy in the presence of covariate collinearity".** In: *Journal of Statistical Planning and Inference* 211, pp. 287–314.

- Lachin, John M (1988). **"Statistical properties of randomization in clinical trials".** In: *Controlled clinical trials* 9.4, pp. 289–311.
- Li, Xinran, Peng Ding, and Donald B Rubin (2018). **"Asymptotic theory of rerandomization in treatment–control experiments".** In: *Proceedings of the National Academy of Sciences* 115.37, pp. 9157–9162.
- — (2020). **"Rerandomization in 2^K factorial experiments".** In.
- Lin, Yunzhi, Ming Zhu, and Zheng Su (2015). **"The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials".** In: *Contemporary clinical trials* 45, pp. 21–25.
- Liu, Zhaoyang et al. (2023). **"Bayesian Criterion for Re-randomization".** In: *arXiv preprint arXiv:2303.07904*.
- Lu, Xin et al. (2023). **"Design-based theory for cluster rerandomization".** In: *Biometrika* 110.2, pp. 467–483.
- Morgan, Kari Lock and Donald B Rubin (2012). **"Rerandomization to improve covariate balance in experiments".** In.

Senn, SJ (1989). **"Covariate imbalance and random allocation in clinical trials".** In: *Statistics in medicine* 8.4, pp. 467–475.

Zhang, Hengtao, Guosheng Yin, and Donald B Rubin (2023). **"PCA rerandomization".** In: *Canadian Journal of Statistics.*

Zhou, Quan et al. (2018). **"Sequential rerandomization".** In: *Biometrika* 105.3, pp. 745–752.