Correlated Cluster-Based Randomized Experiments: Robust Variance Minimization*

Ozan Candogan[1], Chen Chen[2], and Rad Niazadeh[1]

[1]The University of Chicago Booth School of Business
[2]New York University Shanghai
ozan.candogan@chicagobooth.edu, cc8029@nyu.edu, rad.niazadeh@chicagobooth.edu

First version: August 16, 2021
This version: March 22, 2023

Abstract

Experimentation is prevalent in online marketplaces and social networks to assess the effectiveness of new market intervention. To mitigate the interference among users in an experiment, a common practice is to use a cluster-based experiment, where the designer partitions the market into loosely connected clusters and assigns all users in the same cluster to the same variant (treatment or control). Given the experiment, we assume an unbiased Horvitz–Thompson estimator is used to estimate the total market effect of the treatment. We consider the optimization problem of choosing (correlated) randomized assignments of clusters to treatment and control to minimize the worst-case variance of the estimator under a constraint that the marginal assignment probability is $q \in (0, 1)$ for all clusters. This problem can be formulated as a linear program where both the number of decision variables and constraints are exponential in the number of clusters—and hence is generally computationally intractable.

We develop a family of practical experiments that we refer to as *independent block randomization (IBR)* experiments. Such an experiment partitions clusters into blocks so that each block contains clusters of similar size. It then treats a fraction q of the clusters in each block (chosen uniformly at random) and does so independently across blocks. The optimal cluster partition can be obtained in a tractable way using dynamic programming. We show that these policies are asymptotically optimal when the number of clusters grows large and no cluster size dominates the rest. In the special case where cluster sizes take values in a finite set and the number of clusters of each size is a fixed proportion of the total number of clusters, the loss is only a constant that is independent of the number of clusters. Beyond the asymptotic regime, we show that the IBR experiment has a good approximation for any problem instance when q is not very tiny. We also examine the performance of the IBR experiments on data-driven numerical examples, including examples based on Airbnb and Facebook data.

Subject classifications: Variance minimization, robust optimization, experimental design, cluster-based randomization, approximation algorithms, asymptotic optimality.

*An earlier version of this paper was circulated under the title "Near-Optimal Experimental Design for Networks: Independent Block Randomization."

1 Introduction

Experimental design is a celebrated branch of statistics—rooted in the pioneering work of Fisher in the 1920s and 1930s (Fisher 1935). In recent years, thanks to the rapid decrease in the cost of conducting experiments in online platforms, experimental design has become a prevalent tool for improving the operations of online marketplaces and social networks. These platforms often conduct binary experiments, also known as A/B testing, before launching a new feature or introducing a market intervention, as they strive to make data-driven product decisions. To do this, the experiment exposes a (randomized) group of targeted users to the new feature or, equivalently, assigns each user to either the *treatment* or the *control* group. The platform then uses the resulting outcomes to estimate the new feature's total market effect, i.e., the difference in total user outcomes if the feature is introduced to the entire market. Accurately estimating this quantity enables the platform to decide whether to deploy this new feature informatively.[1]

The aforementioned platforms often exhibit complex network effects. Consequently, unless designed carefully, the experiments could suffer from *interference*, where one user's assignment to the treatment or control affects another user's outcome (or behavior). For example, passengers in a ride-sharing platform share the same supply of drivers; hence, enabling prime-time subsidies for passengers in one neighborhood can impact the service experienced by passengers in nearby neighborhoods (Chamandy 2016). Similarly, advertisers in an ad-exchange platform might compete in the same publisher's auction (Basse et al. 2016, Barajas et al. 2016), connected users of a social network might be involved in the same daily activities (Eckles et al. 2017), and hosts in an online hospitality platform (such as Airbnb) might share the same pool of guests (Holtz et al. 2020, Cui et al. 2020). In all of these examples, a user's response to treatment may contaminate the outcomes of other users, thereby resulting in bias or inaccuracy when one estimates the total market effect.

To alleviate such interference, a common practice (e.g., Chapter 22 of Kohavi et al. 2020) is to partition the market into almost disconnected clusters (i.e., groups of users) that exhibit only a minor amount of interference with each other (Eckles et al. 2017, Koutra 2017). For example, riders or hosts from distant neighborhoods are unlikely to interfere with each other; advertisers of totally different products may have different potential publishers; and finally, users in a social network usually form clusters based on their geography, interests, and beliefs. Given these clusters, the platform runs a "cluster-based randomized experiment." Specifically, the platform assigns *all*

[1]When a new feature is rolled out, it often impacts all market participants (see, e.g., Chamandy 2016), which motivates our focus on the total market effect. A complementary research direction, which is outside the scope of the present work, involves personalized feature deployments and adapting the design of experiments accordingly.

of the users in the same cluster to the same treatment or control variant, in order to (hopefully) remove much of the interference, and hence have a relatively unbiased estimation of the total market effect.[2] While such a cluster-based assignment reduces interference, it comes at a cost: since after clustering there are effectively fewer experimental units, the variance can increase, especially if the designer assigns different clusters independently to treatment or control. Thus, the challenge becomes choosing (correlated) randomized assignments across clusters to obtain a lower variance, and understanding the structure of the "optimal correlation" between the binary assignments of different clusters.

1.1 Our Contributions

To address the above challenge, we focus on an ideal setting where the market is partitioned into (heterogeneous) clusters that do not connect/interfere with each other.[3] Such an ideal setting can hold either naturally (e.g., when the market is clustered at the city level in applications such as ride-sharing or online hospitality platforms) or when the designer trusts that clusters have mitigated the interference to an acceptable extent and would like to ignore the remaining inter-cluster interference.

Given the disjoint clusters, the decision maker chooses a random assignment of each cluster to either treatment or control. We assume that the marginal assignment probability is $q \in \left(0, \frac{1}{2}\right]$ (this is without loss of generality by Remark 2.1) for all clusters. Then, she uses the Horvitz–Thompson unbiased estimator (Horvitz and Thompson 1952) to estimate the total market effect given realized outcomes. The objective is to design an optimal joint distribution of assignments under the marginal assignment probability constraint, so as to minimize the variance of this estimator. The variance depends on two things: the joint distribution of assignments and the cluster-level potential outcomes. Since the potential outcomes are uncertain, we formulate the problem as a robust optimization problem against the worst-case values (also known as adversarial values) of the unknown potential outcomes, whose uncertainty sets are nonnegative intervals.

In Section 2, we formally formulate the above optimization problem. Specifically, we show that the experimental design problem is equivalent to a more abstract (robust) optimization problem of *designing (negative) correlation* among n Bernoulli random variables, which could be of independent interest. Each Bernoulli random variable has a marginal success probability q (which corresponds to the probability of being assigned to treatment) and the objective is to minimize the worst-

[2]See Section 1.1 of Wager and Xu (2021) for a related discussion.

[3]This is the same as the "SUTVA (Stable Unit Treatment Value Assumption) for clusters" assumption in, e.g., Hudgens and Halloran (2008), Zigler and Papadogeorgou (2021), and Karrer et al. (2020).

case variance of a weighted sum of these n Bernoulli variables. Here, the weights are selected adversarially from bounded nonnegative intervals.

As a first step to solve the above optimal cluster-based randomized experimental design problem, we focus on the special case where cluster sizes are identical, and we illustrate the *price of independence*. Specifically, we show that randomly assigning a fraction q of the clusters to treatment is optimal (Proposition 3.1), whereas simply assigning treatment to each cluster independently is only a 4-approximation. In general, the problem of obtaining the optimal correlation under the worst-case vector of potential outcomes can be formulated as a linear program. However, this problem is generally intractable, as both the number of decision variables and the number of constraints are exponential in the number of clusters. Moreover, even if we could solve for the optimal experiment, it turns out that it has another subtle drawback: it often involves complicated correlation structures in the assignments of treatments, making it possibly difficult to implement and interpret in practice. Motivated by this, we ask the following natural research question:

> *Can we design simple, computationally efficient, and interpretable correlated cluster-based randomized experiments that are approximately or asymptotically optimal?*

As our main technical contribution, we answer the above question in the affirmative. In particular, inspired by the special case of the optimal cluster-based assignment problem with identical cluster sizes, we develop a family of practical experiments that we refer to as *independent block randomization (IBR)* experiments. Specifically, we partition clusters into blocks so that each block contains clusters of (approximately) similar size. We then assign the treatment variant to a fraction q of the clusters in each block that are selected uniformly at random, and do so independently across blocks. Recall that doing so yields the optimal experiment in the aforementioned special case, but this is not necessarily the case in general. In fact, the suboptimality of the IBR experiments originates from two sources: (a) the loss due to the independence of assignments among different blocks, and (b) the loss from ignoring the cluster size differences within a block. The key idea behind our policies is to partition clusters into blocks in a way that makes these losses as small as possible, and our analysis relies on showing that these losses can indeed be substantially reduced through careful choices of the partitions.

Since the blocks are treated independently, the worst-case variance of an IBR experiment is the sum of the worst-case variances for each block. We provide a full characterization of the worst-case potential outcomes (and hence the worst-case variance) within a block. This characterization gives us a handle to analyze the variance of this family of experiments. Specifically, it enables us to solve for the optimal partitioning of the clusters into blocks in a way that minimizes the variance, through the solution of a simple dynamic program. We then focus on analyzing the performance gap between the optimal IBR experiment and the optimal experiment for the cluster-based assignment problem. We demonstrate and prove the following results:

1. We first (in Section 4.1) establish upper-bounds on the approximation ratio of the optimal IBR experiment for any problem instance with marginal assignment probability q. Our upper-bounds depend on q and lead to acceptable constants when q is not very small (for example, the approximation ratio is $\frac{7}{3}$ for $q = \frac{1}{2}$, 2 for $q = \frac{1}{3}$, $\frac{7}{3}$ for $q = \frac{1}{4}$, and $\frac{12}{5}$ for $q = \frac{1}{5}$). To obtain this result, we analyze a subfamily of simpler k-partition IBR experiments as described in Section 4.1.2. The optimal choice of k in our analysis (which depends on the value of q) provides the desired approximation ratio.

2. We then (in Section 4.2) focus on asymptotic analysis, and show that the IBR experiment with optimal partition is asymptotically optimal for any marginal assignment probability q in the regime with many clusters and when no cluster dominates the rest (asymptotically) in terms of size. We also provide an example that shows that our "no dominating cluster" condition is necessary for any IBR experiment to be asymptotically optimal.

3. In Section 4.2.1, for some special cases of the asymptotic regime, we also obtain stronger results with a more careful analysis of the performance loss. Specifically, when the cluster sizes take values in a finite set that does not change as the problem scales, we show that the simple IBR experiment that places all the clusters of exactly the same size in the same block only increases the worst-case variance by an additive $O(\sqrt{n})$ term compared to the optimal experiment.[4] If, in addition, the number of clusters of each size is a fixed proportion of the total number of clusters, this simple IBR experiment only increases the worst-case variance by an additive $O(1)$ term, i.e., a constant amount.

4. Next, in Section 4.2.2, we introduce a simple member of the IBR family, which we refer to as the *logarithmic-partitioning IBR* experiment. This simple experiment partitions the clusters into blocks by making sure that the ratio of the largest to the smallest cluster sizes in each block is upper-bounded by a given fixed constant $\eta > 1$. We then show that setting η appropriately is sufficient for such an IBR experiment to be asymptotically optimal, albeit at a slower convergence rate compared to the optimal IBR experiment. Notably, this logarithmic-

[4] On the other hand, the worst-case variance of an optimal experiment scales linear in n.

partitioning IBR experiment requires no explicit optimization to find its blocks, which is in contrast to the DP-based optimal IBR experiment.

Hence, our IBR experiments (under a careful design) are near-optimal with respect to minimizing the variance of the Horvitz–Thompson estimator under an adversarial model. More sophisticated designs may have additional benefits, but this improvement is provably marginal.

To complement our theoretical results, in Section 5 and Appendix E, we examine the performance of our IBR experiments on both a synthetic example and data-driven examples based on Airbnb and Facebook data. We demonstrate that our (optimal) IBR experiment performs substantially better than performance guarantees provided by our theoretical results in realistic instances, and (i) decreases the variance substantially relative to independent cluster-based randomization, and (ii) improves upon other heuristic designs—both on average and in the worst case.

1.2 Related Literature

We provide a brief summary of some other related work here, and leave a more comprehensive review of further related literature to Appendix A.

Experiments in Networks and Online Platforms A number of recent papers have studied experimental design with interference in social network settings, e.g., Ugander et al. (2013), Eckles et al. (2017), Aronow and Samii (2017), and Ugander and Yin (2020), and some of these work propose a graph cluster-based randomization. We focus on an ideal case where the network can be partitioned into disjoint clusters (perhaps after ignoring only a small number of connections) and the decision maker is using cluster-based experiments. Also related to us, Pouget-Abadie et al. (2019) introduce a novel correlation clustering objective to extract clusters for cluster-based randomized experiments, while they fix the joint treatment assignments to be completely random. In contrast, we assume the clusters are given exogenously and study the optimal correlation between randomized assignments to minimize the variance of our particular estimator.

Interference in experiments for online markets have also been studied recently, e.g., Johari et al. (2020), Wager and Xu (2021), Bright et al. (2022), and Li et al. (2023). These work often assume a structural model of the marketplace, which induces an interference structure that plays a pivotal role in the design and analysis of the experiments. This is in contrast to our model, where we focus on a robust and efficient experiment that can withstand the worst-case scenario for potential outcomes (with minimal assumption on those, e.g., non-negativity and bounded range).

Robust Design Framework We adopt a robust design approach to experimental design (Berger 2013, Chapter 5) and, in particular, study the problem of minimizing the variance of the estimator against the worst-case value of potential outcomes. Several recent works also use a similar approach. For example, Bojinov et al. (2020) study a switchback experimental design problem. The authors restrict attention to experiments that partition a finite-time horizon into slots, and assign treatment or control variants independently to each slot. Under this restriction, the worst-case potential outcomes take the same extreme point of the uncertainty set, regardless of the experiment. Our problem allows for general joint assignment distributions. The worst-case potential outcomes depend on the specific correlation of the assignments, and thus are experiment-dependent. Harshaw et al. (2019) consider a similar robust design problem, where potential outcomes are assumed to belong to an ℓ_2-ball, and each unit has a covariate that can predict the potential outcomes. A decision maker solves for an optimal experiment to trade off between covariate balancing and robustness. They show that the problem is equivalent to aligning eigenvectors of the resulting correlation matrix in desired directions, and they develop a randomized experiment based on the Gram–Schmidt walk algorithm. Our work considers very different uncertainty sets for potential outcomes. Beyond that, we also differ in other modeling aspects. Specifically, Harshaw et al. (2019) consider a linear relation between potential outcomes and covariates. In constrst, our model does not use covariate information; we only need to know the range of potential outcomes, possibly inferred from covariate information in certain contexts.

1.3 Notation and Terminology

For any two integers $a, b \in \mathbb{N}$ with $a \leq b$, we let $[a:b] = \{a, a+1, \ldots, b-1, b\}$ denote a sequence of integers starting from a and ending with b and we denote $[n] = [1:n]$ for any $n \in \mathbb{N}_+$. For a subset $S \subseteq [n]$, we let $S^c = [n] \setminus S$ denote its complement. For any nonnegative real number $x \in \mathbb{R}_+$, we let $\lfloor x \rfloor \in \mathbb{N}$ denote the floor of x, which is the greatest integer less than or equal to x; and we let $\lceil x \rceil \in \mathbb{N}$ denote the ceiling of x, which is the least integer greater than or equal to x. Finally, we let $\Re$ denote the set of correlation matrices, i.e., matrices that are positive semidefinite with diagonal entries equal to one. The size of the correlation matrices will be clear from the context.

2 Robust Correlation Design for Binary Random Variables

In this section, we first introduce a generic robust optimization problem of *designing (negative) correlation* among n Bernoulli random variables, and we defer its motivation to Section 2.1. Suppose we have n binary random variables $\mathbf{Z} \triangleq \big(Z_i\big)_{i=1}^n \in \big\{0,1\big\}^n$, each with a given marginal distribution Bernoulli(q) for some known value $q \in \big(0, \frac{1}{2}\big]$; however, their correlation matrix needs to be designed. We let $\mathcal{P}_q$ denote the set of all possible joint distributions of these binary random variables $(Z_i)_{i=1}^n$ and $\Sigma(P)$ denote the correlation matrix of $(Z_i)_{i=1}^n$ under the joint distribution $P \in \mathcal{P}_q$. Throughout the paper, we study the following robust correlation design problem:

$$V^{\mathrm{OPT}} = \min_{P \in \mathcal{P}_q} \max_{y \in \times_{i\in[n]}[0,w_i]} y^{\mathrm{T}} \Sigma\big(P\big) y, \tag{1}$$

where a decision maker decides on the correlation matrix of these n binary random variables to minimize the variance of a linear function $y^{\mathrm{T}}\mathbf{Z}$. We further assume that the coefficient vector $y \in \mathbb{R}^n$ is chosen adversarially by nature from the uncertainty set $\times_{i\in[n]}[0, w_i]$, i.e., the cartesian product of the intervals $[0, w_i]$.

In Section 2.1, we motivate this problem as designing the optimal correlation between the assignments of a cluster-based randomized experiment. Specifically, we interpret Z_i as an indicator for the binary assignment of cluster i to either treatment or control, and the joint distribution as a randomized (binary) experiment. In the rest of the paper, we assume that $q \leq \frac{1}{2}$, which is without loss of generality by Remark 2.1.

Remark 2.1. Assuming that the marginal probability q is in $\big(0, \frac{1}{2}\big]$ is without loss of generality. To see this, note that $\mathbb{C}\mathrm{orr}(Z_i, Z_k) = \mathbb{C}\mathrm{orr}\big(1 - Z_i, 1 - Z_k\big)$ for any $i, k \in [n]$; thus, the correlation matrix of $(Z_i)_{i=1}^n$ is the same as that of $(1 - Z_i)_{i=1}^n$. As a result, the correlation design problem (1) with the marginal probability q is the same as that with the marginal probability $1 - q$.

2.1 Optimal Correlation Design for Cluster-Based Randomized Experiments

In this section, we introduce the problem of designing an optimal cluster-based randomized experiment against the worst-case values of potential outcomes, which is our main motivation for studying the min-max optimization (1).

Model Consider a decision maker designing a randomized binary experiment over n disjoint clusters of users as experimental units.[5] . We further restrict our attention to the class of "cluster-based randomized experiments," i.e., picking a (possibly correlated) randomized assignment of each cluster to either of the two possible variants: treatment (the variant "1") or control (the variant "0"). For each cluster $i \in [n]$, we let Z_i be a binary random variable such that $Z_i = 1$ if all users in

[5] We assume clusters are disjoint, possibly after ignoring a small number of edges in the given interference network.

cluster i are assigned to treatment and $Z_i = 0$ if all users in cluster i are assigned to control. A cluster-based randomized assignment with a marginal assignment probability q is specified by a joint distribution $\mathbb{P}[\cdot]$ of these Bernoulli random variables Z_i such that $\mathbb{P}[Z_i = 1] = q$ for all clusters. We assume $q \in \left(0, \frac{1}{2}\right]$ by labeling the variant with a smaller assignment probability as treatment.

For a cluster $i \in [n]$, we let $y_{i1} \in \mathbb{R}$ (and $y_{i0} \in \mathbb{R}$, respectively) be the potential outcome of the cluster (which is the aggregate of the potential outcomes of all users in the cluster) when the cluster receives the treatment variant (the control variant, respectively). Clearly, only one of the potential outcomes y_{i1} and y_{i0} is observed for any cluster i under any assignment.

Objective In applications of binary experimental design (or A/B testing), to make an informed choice between the two variants for all users, the decision maker would like to estimate the *total market effect* τ, which is the difference between the sum of the outcomes when all users receive the treatment and when all users receive the control, i.e.,

$$\tau \triangleq \sum_{i=1}^{n} y_{i1} - \sum_{i=1}^{n} y_{i0} = y_1^{\mathrm{T}}\mathbf{1} - y_0^{\mathrm{T}}\mathbf{1},$$

where $y_1 = (y_{i1})_{i\in[n]}$ and $y_0 = (y_{i0})_{i\in[n]}$ are the concatenations of the potential outcomes and $\mathbf{1}$ is a vector whose entries are all one. The total market effect τ is qualitatively equivalent to the (perhaps more commonly used) average treatment effect (ATE), which is defined as τ divided by the total number of users. Here, we simply focus on the total market effect for a more succinct mathematical exposition later on (as we do not need to carry over the constant normalization throughout).

We focus on the celebrated Horvitz–Thompson unbiased estimator $\hat{\tau}$ (Horvitz and Thompson 1952) to estimate the total market effect, expressed as

$$\hat{\tau} \triangleq \sum_{i\in[n]} \frac{y_{i1}Z_i}{q} - \sum_{i\in[n]} \frac{y_{i0}(1-Z_i)}{1-q} = \frac{y_1^{\mathrm{T}}\mathbf{Z}}{q} - \frac{y_0^{\mathrm{T}}(\mathbf{1}-\mathbf{Z})}{1-q} = y^{\mathrm{T}}\mathbf{Z} - \frac{y_0^{\mathrm{T}}\mathbf{1}}{1-q}, \tag{2}$$

where $y_i \triangleq \frac{y_{i1}}{q} + \frac{y_{i0}}{1-q}$ is the weighted sum of the treatment and control potential outcomes and $y = (y_i)_{i\in[n]}$ denotes their concatenation. By the linearity of expectations, $\mathbb{E}\big[\hat{\tau}\big] = \tau$; thus, the Horvitz–Thompson estimator is indeed unbiased. Given an unbiased estimator, the objective of the decision maker is to design an experiment, i.e., a joint distribution for assigning treatment and control across clusters, to minimize the variance of the estimator, i.e.,

$$\mathbb{V}\mathrm{ar}\left[\hat{\tau}\right] = \mathbb{E}\left[\left(\hat{\tau} - \mathbb{E}[\hat{\tau}]\right)^2\right] = \mathbb{E}\left[\left(\hat{\tau} - \tau\right)^2\right].$$

Note that since the Horvitz–Thompson estimator is unbiased, minimizing the variance is equivalent to minimizing the mean-squared error.

Worst-Case Potential Outcomes and Min-Max Optimization The variance of the Horvitz–Thompson estimator $\hat{\tau}$ depends both on the assignment distribution (i.e., the experiment) and on the value of unknown potential outcomes; hence, we follow a robust design approach. Specifically, we aim to design an experiment that minimizes the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$ of the above estimator against an adversarial selection of the potential outcomes – which basically corresponds to the worst-case potential outcomes y_{i1} and y_{i0} for all $i \in [n]$ for a given experiment. We impose the following assumption on the potential outcomes.

Assumption 2.1 (*Uncertainty Sets of Potential Outcomes*)**.** The potential outcomes are deterministic, nonnegative, and bounded from above; without loss of generality, we assume that $y_{i1} \in [0, w_{i1}]$ and $y_{i0} \in [0, w_{i0}]$ for all clusters $i \in [n]$, where w_{i1} and w_{i0} are a priori known constants.

In many applications, it is reasonable to assume that the potential outcomes are nonnegative;[6] however, our methodology in this paper is general and can be applied to other outcome ranges by proper modifications. We comment further on this in Remark 2.3. The upper bounds w_{i1} and w_{i0} of the uncertainty sets can vary across clusters and treatment/control variants, and can incorporate any prior information on the potential outcomes; e.g., w_{i1} and w_{i0} can be inferred from cluster-level covariates and other side information in practice.

We now formulate the min-max optimization problem of solving the optimal experiment. By (2), we can express the variance of the Horvitz–Thompson estimator as

$$\mathbb{V}\mathrm{ar}[\hat{\tau}] = y^{\mathrm{T}}\mathbb{C}\mathrm{ov}[\mathbf{Z}]y = \tilde{y}^{\mathrm{T}}\Sigma\tilde{y}\,, \tag{3}$$

where $\mathbb{C}\mathrm{ov}[\mathbf{Z}]$ and Σ are the covariance and correlation matrices of the binary random vector $\mathbf{Z}$, respectively, and $\tilde{y} = (\tilde{y}_i)_{i\in[n]}$ with $\tilde{y}_i = \sqrt{q(1-q)} \cdot y_i = \sqrt{q(1-q)} \cdot \left(\frac{y_{i1}}{q} + \frac{y_{i0}}{1-q}\right)$. By Assumption 2.1, we have $\tilde{y}_i \in [0, w_i]$ for each $i \in [n]$, with $w_i \triangleq \sqrt{q(1-q)} \cdot \left(\frac{w_{i1}}{q} + \frac{w_{i0}}{1-q}\right)$. Finally, since $\mathbb{V}\mathrm{ar}[\hat{\tau}]$ is a quadratic convex function of $\tilde{y}$ by (3), in the worst case, each $\tilde{y}_i$ takes a value of either 0 or w_i, which corresponds to either $y_{i1} = y_{i0} = 0$, or $y_{i1} = w_{i1}$ and $y_{i0} = w_{i0}$. We summarize these in Lemma 2.1.

[6]For example, the number of completed rides during a certain time period by a ride-sharing platform, the number of clicks in an online advertisement setting, and the amount of revenue created from a certain marketplace intervention, etc., all need to be nonnegative.

Lemma 2.1 (*Structure of the Worst-Case Potential Outcomes*). *With any cluster-based randomized experiment, the worst-case potential outcome is such that for any cluster $i \in [n]$, either $y_{i1} = y_{i0} = 0$, or $y_{i1} = w_{i1}$ and $y_{i0} = w_{i0}$. Moreover, let Σ be the correlation matrix of the assignments $(Z_i)_{i\in[n]}$; the corresponding variance of the Horvitz–Thompson estimator is*

$$\mathbb{V}\mathrm{ar}[\hat{\tau}] = \tilde{y}^{\mathrm{T}} \Sigma \tilde{y},$$

where $\tilde{y}_i = \sqrt{q(1-q)} \cdot \left(\frac{y_{i1}}{q} + \frac{y_{i0}}{1-q}\right) \in [0, w_i]$ with $w_i = \sqrt{q(1-q)} \cdot \left(\frac{w_{i1}}{q} + \frac{w_{i0}}{1-q}\right)$ for each cluster $i \in [n]$, and $\tilde{y} = (\tilde{y}_i)_{i\in[n]} \in \mathbb{R}^n$ is the concatenation of these values.

By Lemma 2.1, the problem of finding the optimal experiment for minimizing the variance against the adversarial selection of potential outcomes is exactly (1) and the worst-case variance of an optimal experiment is equal to V^{OPT} given the range w_i of the uncertainty sets. We thus have reduced the problem of finding the "robust" optimal cluster-based randomized experiment to (1).

For ease of exposition, and motivated by this application, in the remainder of the paper we refer to the min-max optimization in (1) using the cluster-based randomized experiment terminology. Specifically, we refer to index i as an index of the clusters, an upper bound w_i of the uncertainty set as the size of cluster i, a random variable Z_i as a treatment assignment to cluster i, and a joint assignment distribution of $(Z_i)_{i=1}^n$ as an experiment.[7]

2.2 Hardness of Solving the Min-max Optimization Problem (1)

Since all potential outcomes are nonnegative by Assumption 2.1, achieving a small objective value in (1) necessitates having a correlation matrix $\Sigma(P)$ with large (in absolute terms) negative off-diagonal entries. Thus, intuitively, (1) can be viewed as a problem of designing *optimal negative correlation* among the assignments. Given that the problem of achieving large negative correlations between pairs of random variables (through the appropriate choice of some decision variables) is quite natural, we suspect that our formulation and approach could be of interest in other settings as well. Note that the min-max problem (1) is challenging to solve exactly, because:

(i) First, the inner problem is to maximize a quadratic convex function. Since the objective is not concave, off-the-shelf optimization algorithms do not guarantee achieving an optimal solution.

[7] The size of a cluster should be more precisely referred to as the size of the uncertainty set. Usually, a cluster with a larger physical size (i.e., number of users) tends to have a more extensive uncertainty set. On the other hand, even when clusters have equal physical sizes, they can still be heterogeneous in other important covariates and hence can have different ranges of potential outcomes, in which case our framework still applies. Here, we simply refer to the upper bound of the uncertainty set as the size of a cluster for ease of exposition, at the expense of a little ambiguity.

Moreover, due to convexity, the maximum is always achieved at an extreme point of the feasible region satisfying $y_i \in \{0, w_i\}$ for all $i \in [n]$. One way to solve this problem is to evaluate the objective at the extreme points, which is computationally difficult due to the exponential number of extreme points (in the number of clusters).

(ii) Second, the outer problem involves minimization over the joint distribution of binary assignments, and the number of decision variables is exponential in the number of clusters as well. It may be possible to develop approximation algorithms that, e.g., rely on a semidefinite programming relaxation of the inner problem, and a relaxation of the outer problem to allow for any correlation matrix—not necessarily only those achievable by a binary random assignment. But it remains unclear how to efficiently compute a joint binary assignment distribution from such a correlation matrix, even when the correlation matrix is indeed feasible.[8]

(iii) Third, (1) can be formulated as a linear program (see (6) in Appendix B.3). However, this program has an exponential number of decision variables and constraints in the number of clusters: we have the constraints defining the feasible set $\mathcal{P}_q$ plus we have one constraint for each extreme point of the potential outcomes' uncertainty set (to encode the worst-case objective). Hence, solving this linear program directly is not computationally tractable.

Motivated by these challenges, in Section 3, we consider a family of practical experiments that we refer to as independent block randomization experiments, and we show that (i) they are easy to compute and interpret, and (ii) they admit provable performance guarantees. Before introducing this family formally, we conclude this section with two remarks: one on the min-max formulation of (1), and the other one on the nonnegative potential outcomes assumption.

Remark 2.2 (*The Min-Max Formulation*). Rüschendorf and Uckelmann (2002) and Section 3.6 of Rachev and Rüschendorf (1998) study a technically related problem to our robust variance minimization problem (1): given random variables with fixed marginal distributions, find a joint distribution for these random variables to minimize the variance of a linear combination of these variables (with known linear coefficients). The min-max formulation of our problem, in which the linear coefficients y are adversially chosen given a randomized experiment (rather than being fixed and known a priori), renders fundamental differences from the problem described above. Specifically, suppose that the marginal assignment probability q is equal to $\frac{1}{2}$. If all the coefficients y are known in advance, the optimal experimental design problem simply reduces to the classic balanced cut

[8] For the case of $q = \frac{1}{2}$, given a correlation matrix $\Sigma \in \Re$, one can obtain a heuristic random assignment based on the standard random hyperplane rounding ideas (Chapter 6 of Williamson and Shmoys 2011). This random assignment has a correlation matrix $\frac{2}{\pi}\arcsin(\Sigma) \succeq \frac{2}{\pi}\Sigma$, and can increase the worst-case variance considerably compared to the worst-case variance associated with the correlation matrix Σ.

of y—which is NP-hard in general (Spielman and Teng 2004); see Appendix B.1 for more details. Our problem does not have this simple structure; actually, the adversarial selection of coefficients y had made our problem more complicated than the one in Rüschendorf and Uckelmann (2002). To minimize the worst-case variance, we need to carefully design the negative correlation among clusters, rather than simply partitioning the clusters in a balanced way as in Appendix B.1.

Remark 2.3 (*Alternative Uncertainty Sets*)**.** In this paper, we restrict attention to settings with non-negative potential outcomes (as captured by Assumption 2.1). Importantly, our approach is general and can be applied to other ranges of potential outcomes as well. It turns out that the choice of the range of potential outcomes qualitatively impacts the structure of the optimal experimental design. For instance, for the special case where the uncertainty set of each cluster i's potential outcomes is symmetric around zero, i.e., $y_{i1} \in [-w_{i1}, w_{i1}]$ and $y_{i0} \in [-w_{i0}, w_{i0}]$, an optimal experiment solving (1) simply assigns each cluster independently to treatment with probability q (see Appendix B.2). Interestingly, this simple experimental design seems to be optimal in very restrictive settings. When the potential outcomes do not belong to an interval symmetric around zero, e.g., as in the setting considered in this section, more intricate designs are needed to ensure low variance in the worst case. We revisit this point and illustrate the suboptimality of independent assignments in the absence of symmetry in the next section as well as Section 5.

3 Independent Block Randomization

In this section, we introduce our main family of experiments that attain near-optimal solutions to (1). To provide insights into the design of the experiments in this family, we first (in Section 3.1) discuss the *price of independence*, i.e., the variance gap between the optimal solution to (1) and the random assignment that treats clusters independently. Subsequently, in Section 3.2, we formally define our proposed family of experiments, which are defined through a partition of clusters. We also provide an approach to efficiently computing the optimal partition. Finally, in Appendix F, we discuss a way to construct confidence intervals for an IBR experiment.

3.1 Price of Independence: Optimal Experiment with Equal-Sized Clusters

Before we introduce the family of IBR experiments, as a warm-up, we study the case where cluster sizes $(w_i)_{i=1}^n$ are equal. In this case, Proposition 3.1 shows that the optimal experiment randomly assigns a fraction q of the clusters to treatment.

Proposition 3.1 (*Optimal Experiments with Equal-Sized Clusters*). *Suppose that the upper bounds of the uncertainty sets $(w_i)_{i=1}^n$ in (1) are equal, and without loss of generality, suppose that $w_i = 1$ for all $i \in [n]$. The optimal experiment solving (1) is as follows:*

1. *When $qn \in \mathbb{N}$: An optimal experiment chooses qn clusters uniformly at random and assigns them to treatment, while assigning the rest to control. The correlation coefficient of any two assignments is $\sigma = -\frac{1}{n-1}$. Furthermore, if n is even, in the worst case, $\frac{n}{2}$ clusters take $y_i = 1$ and the other clusters take $y_i = 0$, and $V^{\text{OPT}} = \frac{1}{4}\frac{n^2}{n-1}$. If n is odd, in the worst case, either $\frac{n+1}{2}$ or $\frac{n-1}{2}$ clusters take $y_i = 1$ and the other clusters take $y_i = 0$, and $V^{\text{OPT}} = \frac{n+1}{4}$.*
2. *When $qn \notin \mathbb{N}$: An optimal experiment, with probability $p \triangleq \lceil qn \rceil - qn$, chooses $\lfloor qn \rfloor$ clusters uniformly at random and assigns them to treatment, and with probability $1-p$, chooses $\lceil qn \rceil$ clusters uniformly at random and assigns them to treatment, and then assigns the rest to control. The correlation coefficient of any two cluster assignments is $\sigma = -\frac{nq(1-q)-p(1-p)}{n(n-1)q(1-q)} \in \left(-\frac{1}{n-1}, 0\right)$. Let $h^* \in \mathbb{N}$ denote the integer closest to $\min\left\{-\frac{1}{2\sigma} + \frac{1}{2}, n\right\}$. In the worst case, h^* clusters take $y_i = 1$ and the rest take $y_i = 0$, and $V^{\text{OPT}} = h^* + h^*(h^*-1)\sigma$. Moreover, $\lim_{n\to\infty} \frac{4V^{\text{OPT}}}{n} = 1$.*

We prove Proposition 3.1 in Appendix B.3. Importantly, this result not only characterizes the optimal experiment in this case, but also shows that there is an inherent gap between the variance of the experiment that treats clusters independently and that of the optimal experiment. Specifically, the worst-case variance of the optimal experiment asymptotically converges to $\frac{n}{4}$, while it is straightforward to see from (1) that the worst-case variance of the independent assignment is exactly equal to n. Hence, there is a multiplicative gap of 4 between the two variances, which characterizes the price of independence for solving (1).

Finally, we provide the optimal experiment when the marginal assignment probability q is equal to $\frac{1}{2}$ and clusters have equal sizes in Corollary 3.2, which follows directly from Proposition 3.1.

Corollary 3.2 (The Case of $q = \frac{1}{2}$). *Suppose that all the cluster sizes are equal with $w_i = 1$ for each $i \in [n]$, and the marginal assignment probability q is equal to $\frac{1}{2}$. The optimal experiment solving (1) is as follows:*

1. *When n is even: The optimal experiment chooses $\frac{n}{2}$ clusters uniformly at random and assigns them to treatment, while assigning the rest to control. The correlation coefficient of any two assignments is $\sigma = -\frac{1}{n-1}$. In the worst case, $\frac{n}{2}$ clusters take $y_i = 1$ and the other $\frac{n}{2}$ clusters take $y_i = 0$, and $V^{\text{OPT}} = \frac{1}{4}\frac{n^2}{n-1}$.*

2. *When n is odd: The optimal experiment chooses $\frac{n+1}{2}$ clusters uniformly at random, assigns them to treatment and the rest to control with probability $\frac{1}{2}$, and vice versa with probability $\frac{1}{2}$. The correlation coefficient of any two assignments is $\sigma = -\frac{1}{n}$. In the worst case, $\frac{n+1}{2}$ clusters take $y_i = 1$ and the other $\frac{n-1}{2}$ clusters take $y_i = 0$, and $V^{\text{OPT}} = \frac{1}{4}\frac{(n+1)^2}{n}$.*

3.2 Optimal Independent Blocks and Dynamic Programming

Proposition 3.1 shows that when clusters have equal sizes, an optimal experiment treats clusters in an identical way, and it minimizes the correlation between assignments to any two clusters. This smallest (negative) correlation is achieved by assigning a fraction q of the clusters to treatment uniformly at random. In general, when cluster sizes are different, although we would like the correlation to be small for any two clusters, an optimal experiment will prioritize the negative correlation between some specific cluster pairs (e.g., when both clusters are large) over some other pairs (when both clusters are relatively small). It may even deliberately introduce positive correlation between some pairs of clusters in order to attain larger negative correlation between other pairs. It is not clear what the optimal correlation among clusters would look like or how to search for it in a computationally efficient way.

Overview of IBR Experiments Inspired by the case with equal cluster sizes, we consider a family of simple experiments that we refer to as *independent block randomization (IBR)* experiments. Specifically, in an IBR experiment, we first sort clusters in decreasing order of size. We then partition them into blocks so that each block contains clusters of similar sizes.[9] We then try to obtain assignments of any two clusters in a block to treatment and control in a way that induces large negative correlation. To do this, we uniformly at random treat a fraction q of the clusters in each block, and do so independently across blocks. Note that Proposition 3.1 implies that the aforementioned assignment attains the largest negative correlation among clusters in a block, ignoring the differences in size. Thus, intuitively, this assignment ensures large negative correlation when the cluster sizes in a block are not too different. The independence of assignments across blocks, on the other hand, comes at the price of no correlation between clusters of different blocks.

A careful design of the blocks mentioned above trades off between: (i) the benefit of a larger negative correlation within a block, and (ii) the price of independence across blocks. Specifically, as a block contains more clusters, more clusters are negatively correlated with each other, but simultaneously, the correlation also becomes weaker (i.e., the magnitude of the correlation becomes

[9] In fact, we sort *before* partitioning precisely because we would like to ensure similar size in the same block.

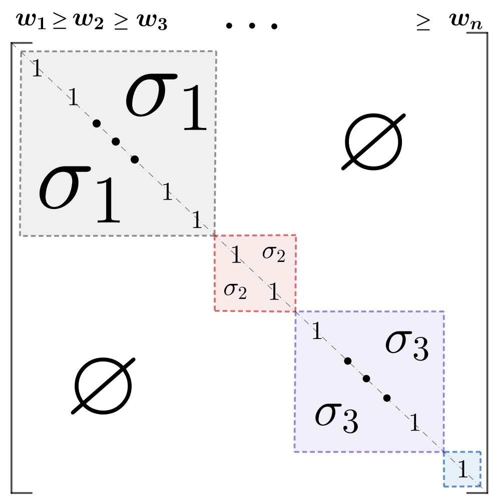


Figure 1: Example of the correlation matrix of an IBR experiment (with four blocks, where the last block is a single cluster). Note that for each block i, $\sigma_i < 0$ and $|\sigma_i|$ decrease as the size of the block increases.

smaller; we prove this in Lemma C.1 of Appendix C). Example 3.1 illustrates this point by focusing on two extreme cases. See also Figure 1 for an illustration of the structure of the correlation matrix of the random assignments generated by an IBR experiment. This structure shows how our IBR experiments control the induced negative correlation between different clusters.

Example 3.1. If there is only one block that contains all the clusters, then the assignments of any two clusters are negatively correlated, but the correlation is only $\Theta\left(-\frac{1}{n}\right)$. This turns out to be optimal when the cluster sizes are equal, by Proposition 3.1. If each block instead contains only two clusters, then in the associated IBR experiment, the correlation between the assignments of these two clusters has the largest absolute value possible (in particular, it is -1 when $q = \frac{1}{2}$). However, in this case, every cluster is only (negatively) correlated with the cluster in the same block, and is independent of the remaining $n-2$ clusters. More generally, as the size of a block increases, the correlation of any two clusters in the block becomes smaller in absolute value by Lemma C.1.

We next show that the optimal partition of clusters can be obtained through the solution of a simple dynamic program. Our approach builds on Lemma 3.3, which characterizes a worst-case potential outcome of a block in an IBR experiment.

Lemma 3.3 (*Worst-Case Potential Outcomes of a Block*). *Consider a block with k clusters sorted in decreasing order of size, i.e., $w_1 \geq w_2 \geq \cdots \geq w_k$. Let $\Sigma \in \Re$ denote the correlation matrix of assignments with off-diagonal entries all equal to σ, and let r be the largest index such that $w_r \geq -2\sigma \sum_{i \leq r-1} w_i$. Let $y_i = w_i$ for all $i \leq r$, and $y_i = 0$ for all $i > r$. Then, $y = \left(y_i\right)_{i\in[k]}$ is a worst-case potential outcome of the block, i.e., it solves $\max_{y_i\in[0,w_i],\,\forall\, i\in[k]}\ y^{\mathrm{T}}\Sigma y$.*

We prove Lemma 3.3 in Appendix B.4. In an IBR experiment, the assignments to blocks are independent; thus, the worst-case variance is simply the sum of worst-case variances of different blocks. This fact and Lemma 3.3 together indicate that we can obtain an optimal cluster partition in polynomial time using dynamic programming.

Specifically, given the sorted list $w_1 \geq w_2 \geq \ldots \geq w_n$, we let $V_k(h)$ denote the continuation worst-case variance when there are $k \in [n]$ remaining clusters $[n-k+1:n] \triangleq \{n-k+1, \ldots, n\}$ to be partitioned and the next block contains $h \in [k]$ clusters. Let $V_k = \min_{h \in [k]} V_k(h)$ be the worst-case variance from the optimal partition of these k clusters. The Bellman equation is

$$V_k = \min_{h \in [k]} V_k(h) = \min_{h \in [k]} g_k(h) + V_{k-h}, \tag{4}$$

where $g_k(h)$ is the worst-case variance of the block that contains the first h of the remaining k clusters. Note that this quantity can be easily computed using Lemma 3.3.[10] In the remainder of the paper, we let $V^{\text{DP}} \triangleq V_n$ denote the worst-case variance of the optimal IBR experiment.

We conclude this section with a remark that compares our independent block randomization experiments with stratified randomization experiments.

Remark 3.1 (*Comparison to Stratified Randomization*). Stratified randomization experiments (Fisher 1935, Higgins et al. 2016) group (or stratify) the experimental units (usually based on their covariates) and then assign treatments to each group independently. We can interpret our independent block randomization experiment as stratification in cluster sizes w_i (which may potentially also incorporate relevant covariant information or other side information). However, the reason for this "size stratification" in our case is very different from conventional stratified randomization. Specifically, we stratify cluster sizes not because we assume clusters of similar size to have similar potential outcomes (which is the usual motivation). Instead, we establish that stratification in cluster sizes allows for reducing the worst-case variance of the estimator (see the analysis in Section 4). Actually, as can be seen from Lemma 3.3, although clusters in the same block have similar sizes, their potential outcomes take distinct values in the worst case. Specifically, some potential outcomes take the largest values w_i, whereas the rest take zero.

[10] As a side note, when the marginal assignment probability $q = \frac{1}{2}$, it is without loss of optimality (see Lemma C.3 in the Appendix) to focus on blocks that contain an even number of clusters, except for the last block that contains the smallest clusters (which has an odd number of clusters when n is odd). This property quarters the computational requirement for solving the DP.

4 Performance Analysis of the IBR Experiments

In this section, we analyze the performance of our IBR experiments. When the cluster sizes are heterogeneous, the suboptimality of an IBR experiment derives from two sources:

(i) Assignments are independent across blocks; hence, we lose the opportunity to introduce negative correlation between clusters of different blocks (which would yield lower total variance in the worst-case outcome).

(ii) Within each block, the cluster sizes are not exactly the same, but we treat the clusters in an identical way.

Our design of IBR experiments tries to mitigate both sources of suboptimality. In fact, our analysis relies on showing that these losses can indeed be substantially reduced through careful choices of the partitions. We consider both an approximation ratio analysis for any problem instance (Section 4.1) and an asymptotic analysis when the number of clusters is large (Section 4.2). In both cases, we show the worst-case variance increases only by a small amount compared to the worst-case variance V^{OPT} of an optimal experiment.

Remark 4.1. Throughout this section, for notational convenience, we assume clusters are sorted in decreasing order of size, i.e., $w_1 \geq w_2 \geq \cdots \geq w_n$.

In the analysis, instead of comparing the performance of IBR experiments to the optimal worst-case variance V^{OPT} directly, we compare it to a lower bound V^{LB} of (1). The lower bound is achieved by relaxing the outer problem to allow for any correlation matrix. More precisely, we have:

$$V^{\mathrm{LB}} \triangleq \min_{\Sigma \in \Re} \max_{y \in \times_{i \in [n]} [0, w_i]} y^{\mathrm{T}} \Sigma y \leq V^{\mathrm{OPT}}. \tag{5}$$

The key difference of this problem from (1) is that in (5), Σ belongs to the set of all correlation matrices $\Re$, and it need not be attained by a joint binary assignment.[11] As a result of this relaxation, we have the following lemma.

Lemma 4.1 (*Relaxation of (1)*)**.** *Suppose that V^{OPT} and V^{LB} are the optimal objective values of the min-max optimizations* (1) *and* (5)*, respectively. Then, we have $V^{\mathrm{LB}} \leq V^{\mathrm{OPT}}$.*

[11] The set of all correlation matrices is equivalent to the set of symmetric positive semidefinite matrices with all diagonal entries being one—simply because any such matrix can be induced by a joint Gaussian distribution. Note that this set is not a polyhedron. On the other hand, the set of correlation matrices $\Sigma(P)$ with a random joint binary assignment $P \in \mathcal{P}_q$ is indeed a polyhedron.

4.1 Approximation Ratio Analysis

In this section, we show that a simple IBR experiment achieves a good approximation ratio for any problem instance if the marginal assignment probability q is not very tiny; this in turn implies that the optimal IBR experiment from solving the DP in (4) achieves a good approximation ratio as well. For example, we show that the optimal IBR experiment has an approximation ratio of $\frac{7}{3}$ for $q = \frac{1}{2}$, 2 for $q = \frac{1}{3}$, $\frac{7}{3}$ for $q = \frac{1}{4}$, and $\frac{12}{5}$ for $q = \frac{1}{5}$.

4.1.1 Preparation: Analysis of Independent Assignments

To start, we first consider a naive experiment that treats every cluster independently with probability q, which is equivalent to having one block for each cluster. The corresponding correlation matrix is $\Sigma = I$, and the worst-case variance is $\sum_{i \in [n]} w_i^2$. We first provide a lower bound on V^{LB} and show that this naive experiment is a 4-approximation.

Lemma 4.2 (*Approximation Ratio of the Independent Assignment*)**.** *The worst-case variance V^{OPT} of an optimal experiment satisfies*

$$\max\left\{w_1^2, \frac{1}{4}\sum_{i \in [n]} w_i^2\right\} \le V^{\text{LB}} \le V^{\text{OPT}} \le \sum_{i \in [n]} w_i^2.$$

We prove Lemma 4.2 in Appendix B.5. The lower bound $\frac{1}{4}\sum_{i \in [n]} w_i^2 \le V^{\text{LB}}$ implies that the aforementioned naive independent assignment is a 4-approximation. Note that the approximation ratio 4 is (asymptotically) tight. To see this, consider a problem instance with n clusters where all cluster sizes are equal to one. By Proposition 3.1, we have $\lim_{n \to \infty} \frac{\sum_{i \in [n]} w_i^2}{V^{\text{OPT}}} = 4$.

4.1.2 Analytical Tool: k-Partition IBR Experiments

We now use Lemma 4.2 as a building block to analyze a more advanced IBR experiment that groups all fixed k clusters together, and then solves for the optimal partition for each group separately in order to find the final blocks. We call such an experiment a *k-partition IBR experiment*, and we let V^k denote the worst-case variance of this experiment.

More specifically, let $N = \left\lceil \frac{n}{k} \right\rceil$ be the number of groups. Every group $h \in [1 : N - 1]$ contains exactly k clusters $(h-1)k + i$ for $i = 1, 2, \ldots, k$, and the last group $h = N$ contains the remaining clusters (hence, it can have fewer than k clusters). Now we solve the DP for each group separately to obtain the optimal partition for the group, and we combine these partitions as the final partition in the k-partition IBR experiment.

q	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9	1/10
k^*	4	3	4	5	6	7	8	9	10
bound on $\frac{V^{k^*}}{V^{\mathrm{LB}}}$	$\frac{7}{3} \approx 2.33$	2	$\frac{7}{3} \approx 2.33$	2.4	2.7	$\frac{20}{7} \approx 2.86$	$\frac{22}{7} \approx 3.14$	$\frac{10}{3} \approx 3.33$	$\frac{65}{18} \approx 3.61$

Table 1: The minimum approximation ratio bound of the k-partition IBR experiments with some specific values of q. Here, $k^* = \operatorname{argmin}_{k \in \mathbb{N}_+} f_q(k)\big(\frac{4}{k} + \frac{k-1}{k}\big)$ is the optimal value of the parameter k, and the bound on $\frac{V^{k^*}}{V^{\mathrm{LB}}}$ (i.e., the third line) refers to $f_q(k^*)\big(\frac{4}{k^*} + \frac{k^*-1}{k^*}\big)$.

We let $f_q(k)$ denote the worst-case variance (i.e., the optimal value of (1)) for the problem instance with k clusters and all cluster sizes equal to one, which is easy to compute by Proposition 3.1. Lemma 4.3 bounds the approximation ratio of the k-partition IBR experiment in terms of $f_q(k)$ for any integer k.

Lemma 4.3 (*Approximation Ratio of the k-Partition IBR Experiment*). *For any integer k, the worst-case variance V^k of a k-partition IBR experiment satisfies*

$$\frac{V^k}{V^{\mathrm{OPT}}} \le \frac{V^k}{V^{\mathrm{LB}}} \le f_q(k)\left(\frac{4}{k} + \frac{k-1}{k}\frac{w_1^2}{V^{\mathrm{LB}}}\right) \le f_q(k)\left(\frac{4}{k} + \frac{k-1}{k}\right),$$

where $f_q(k)$ is the optimal value of (1) for the problem instance with k clusters and all cluster sizes equal to one.

We prove Lemma 4.3 in Appendix B.6. Note that when $k = 1$, the k-partition IBR experiment coincides with the naive independent assignment. On the other hand, since $f_q(1) = 1$ for any $q \in (0, 1)$, Lemma 4.3 again shows that the independent assignment is a 4-approximation.

Using Lemma 4.3, for any marginal assignment probability q, we can search over k and minimize $f_q(k)\left(\frac{4}{k} + \frac{k-1}{k}\right)$ to find the k-partition IBR experiment that attains the minimum approximation ratio bound. We illustrate such a minimum approximation ratio and the corresponding value of k in Figure 2, and we provide details for some specific values of q in Table 1. As can be seen from Figure 2, the best k-partition IBR experiment reduces the worst-case variance considerably compared to the independent assignment (whose approximation ratio 4 is tight) at least when the marginal assignment probability q is not that tiny, i.e., for $q \in [0.1, 0.5]$.

We next consider a special case where the marginal assignment probability q equals $\frac{1}{m}$ for some integer m. Interestingly, as can be seen from Figure 2 and Table 1, except when the optimal value of k, denoted by k^*, is 4 for $m = 2$, the optimal value $k^* = m$ for $m \in [3:10]$. On the other hand, by Proposition 3.1, as long as a block contains no more than m clusters, the correlation of assignments between any two clusters in the block is always $-\frac{1}{m-1} = -\frac{q}{1-q}$ under the uniformly random assignment in Proposition 3.1. Thus, with m clusters, it is optimal to put all these clusters into one block and, as a result, the m-partition IBR experiment simply groups all m clusters into one block and, within each block, it selects one cluster uniformly at random and assigns it to treatment. We can view the m-partition IBR experiment as a generalization of the pair-matching experiment (to be discussed in Remark 4.4) from the case of $q = \frac{1}{2}$ to the case of $q = \frac{1}{m}$. Proposition 3.1 and Lemma 4.3 directly imply an approximation ratio guarantee for such a $\frac{1}{q}$-partition IBR experiment, which we state in Corollary 4.4. Note that the approximation ratio guarantee in Corollary 4.4 is only meaningful when m is sufficiently small (and hence q is sufficiently large), because the $\frac{1}{q}$-partition IBR experiment cannot be worse than the independent assignment, which has a tight approximation factor 4.

Corollary 4.4 (Approximation Ratio of the $\frac{1}{q}$-Partition IBR Experiment)**.** *Suppose that the marginal assignment probability q is $\frac{1}{m}$ with some integer m. The m-partition IBR experiment simply groups all m clusters into one block (except for the last block that contains the smallest clusters, which can have fewer than m clusters), and selects one cluster from each block uniformly at random and assigns it to treatment (except for the last block if it is not full). Moreover, if m is odd, we have $\frac{V^m}{V^{\mathrm{LB}}} \leq \min\left\{4, \frac{m+1}{m} + \frac{m^2-1}{4m}\right\}$, and if m is even, $\frac{V^m}{V^{\mathrm{LB}}} \leq \min\left\{4, \frac{m}{m-1} + \frac{m}{4}\right\}$.*

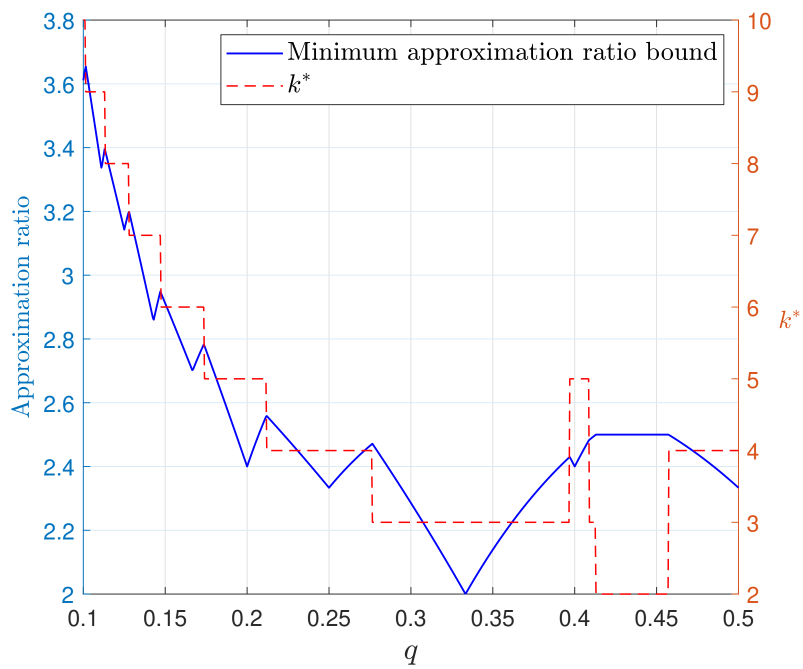


Figure 2: The red dashed line with right y-axis: $k^* = \mathrm{argmin}_{k\in\mathbb{N}_+} f_q(k)\left(\frac{4}{k} + \frac{k-1}{k}\right)$ is the optimal value of the parameter k for the k-partition IBR experiment to attain the minimum approximation ratio bound, with a given marginal assignment probability q. The blue solid line with left y-axis: The minimum approximation ratio bound with a given marginal assignment probability q, which equals $f_q(k^*)\left(\frac{4}{k^*} + \frac{k^*-1}{k^*}\right)$.

Lemma 4.3 directly implies that the optimal IBR experiment admits an improved approximation factor over the independent assignment for any problem instance, at least when the marginal assignment probability is not very tiny; we state the approximation ratio guarantee for the optimal IBR experiment in Theorem 4.5.

Theorem 4.5 (*Approximation Ratio of the Optimal IBR Experiment*)**.** *The worst-case variance of the optimal IBR experiment from solving the DP satisfies*

$$\frac{V^{\mathrm{DP}}}{V^{\mathrm{OPT}}} \le \frac{V^{\mathrm{DP}}}{V^{\mathrm{LB}}} \le \min_{k\in\mathbb{N}_+} \frac{V^k}{V^{\mathrm{LB}}} \le \min_{k\in\mathbb{N}_+} f_q(k)\left(\frac{4}{k}+\frac{k-1}{k}\right).$$

4.1.3 The Case of $q = \frac{1}{2}$

We now turn to the case where the marginal assignment probability q equals $\frac{1}{2}$. By Corollary 3.2, $f_{\frac{1}{2}}(k) = \frac{1}{4}\frac{k^2}{k-1}$ if k is even and $f_{\frac{1}{2}}(k) = \frac{1}{4}\frac{(k+1)^2}{k}$ if k is odd. Thus, by Lemma 4.3, we have $\frac{V^k}{V^{\mathrm{LB}}} \le \frac{k}{k-1} + \frac{k}{4}$ when k is even and $\frac{V^k}{V^{\mathrm{LB}}} \le \frac{(k+1)^2}{k^2} + \frac{(k+1)^2(k-1)}{4k^2}$ when k is odd. Note that when k is odd, since $\frac{(k+1)^2}{k^2} + \frac{(k+1)^2(k-1)}{4k^2} > \frac{k+1}{k} + \frac{k+1}{4}$, the approximation ratio bound with a k-partition IBR experiment is always larger than the approximation ratio bound with a $(k+1)$-partition IBR experiment. Thus, we only need to focus on the case where k is even. In this case, the minimum approximation ratio is attained with $k = 4$, which yields $\frac{V^{k=4}}{V^{\mathrm{OPT}}} \le \frac{7}{3}$. The value of $k = 2$ yields $\frac{V^{k=2}}{V^{\mathrm{OPT}}} \le \frac{5}{2}$, which is slightly larger. We summarize these in Corollary 4.6.

Corollary 4.6 (Approximation Ratio Guarantee with $q = \frac{1}{2}$)**.** *Suppose that the marginal assignment probability q equals $\frac{1}{2}$. The worst-case variance V^k of the k-partition IBR experiment with an even integer k satisfies*

$$\frac{V^k}{V^{\mathrm{OPT}}} \le \frac{V^k}{V^{\mathrm{LB}}} \le \frac{k}{k-1} + \frac{k}{4}\cdot\frac{w_1^2}{V^{\mathrm{LB}}} \le \frac{k}{k-1} + \frac{k}{4}.$$

The minimum approximation ratio bound is attained with $k = 4$, which yields $\frac{V^{k=4}}{V^{\mathrm{OPT}}} \le \frac{7}{3}$; this implies that the optimal IBR experiment is a $\frac{7}{3}$-approximation as well. In addition, the approximation ratio of the 2-partition IBR experiment satisfies $\frac{V^{k=2}}{V^{\mathrm{OPT}}} \le \frac{5}{2}$.

For the case of q equal to $\frac{1}{2}$, as discussed earlier (and detailed in Lemma C.3 of the Appendix), in the optimal partition from solving the DP, all blocks will have an even number of clusters (except for the last block that contains the smallest clusters when n is odd). Thus, to obtain the optimal partition of a set of $k = 4$ clusters (as required by the k-partition IBR experiment), we only need to compare the worst-case variances between two cases: (a) a block that contains all four clusters, and (b) two blocks with one block containing the first two clusters and the second one containing the other two clusters. Thus, the 4-partition IBR experiment has a minimal computational requirement.

We conclude this section with three remarks: one on the theoretical performance of our optimal IBR experiment in Theorem 4.5 versus its numerical performance, one about our analysis for small marginal assignment probability q, and finally one on the theoretical performance of the commonly used pair-matching experiment (and its generalization). We also provide an upper bound on the $f_q(k)$ function in the end (Lemma 4.7), which is useful for the asymptotic analysis in Section 4.2.

Remark 4.2 (*Theoretical versus Numerical Performance Guarantees*). In Theorem 4.5, we provide a theoretical approximation ratio guarantee for the optimal IBR experiment based on analyzing a subfamily of simpler k-partition IBR experiments. On the other hand, one may wonder how tight these analyses are. Take the special case of $q = \frac{1}{2}$ as an example; in Section 5.1, we construct problem instances where the approximation ratio for the optimal IBR experiment is no better than approximately $1.52 < 7/3$, and the approximation ratio for the 4-partition IBR experiment is no better than approximately[12] $1.62 < 7/3$. But we do not have an example where the approximation ratios are substantially closer to 7/3. This gap can be partially attributed to the slack in the analysis of the k-partition IBR experiment and the improvement of the optimal IBR experiment over the k-partition IBR experiment. Furthermore, our numerical study in Section 5 demonstrates that the optimal IBR experiment often performs substantially better than the theoretical performance guarantee in Theorem 4.5. This is illustrated by focusing both on a synthetic example with randomly generated instances and on data-driven examples based on Airbnb and Facebook data.

Remark 4.3 (*Analysis Slack for Small q*). We would like to note that our theoretical analysis is not tight when the marginal assignment probability q is small. By the discussion above Corollary 4.4, it is optimal to let each block contain at least $\left\lfloor \frac{1}{q} \right\rfloor$ clusters. Therefore, when q is small, we want a large k for the k-partition IBR experiment. On the other hand, our analysis for the k-partition IBR experiment is loose when k is large. In particular, we increase all cluster sizes in a group of k clusters to the maximum cluster size of the group, in order to bound the worst-case variance of the k-partition IBR experiment from above (please refer to the proof of Lemma 4.3). This yields the $\frac{k-1}{k}$ term in Lemma 4.3 (note that $f_q(k) = \Theta(k)$ by Proposition 3.1 and Lemma 4.7). Although numerical examples demonstrate that our optimal IBR experiment still reduces the worst-case variance substantially relative to independent randomization when q is small, our current analysis does not support this; we leave further investigations on a refined analysis to future research.

[12]Numerical examples in Section 5.1 also show that the approximation ratio for the optimal IBR experiment is no better than approximately $1.49 < 2$ for $q = \frac{1}{3}$, $1.51 < \frac{7}{3}$ for $q = \frac{1}{4}$, and $1.47 < \frac{12}{5}$ for $q = \frac{1}{5}$, and the approximation ratio for the m-partition IBR experiment with $m = 1/q$ is no better than approximately $1.63 < 2$ for $q = \frac{1}{3}$, $1.55 < \frac{7}{3}$ for $q = \frac{1}{4}$, and $1.53 < \frac{12}{5}$ for $q = \frac{1}{5}$.

Remark 4.4 (*Performance Guarantees for the Pair-Matching Experiment and Its Generation*)**.** The pair-matching experiment – or the pair experiment for short – is a commonly used heuristic (Imai et al. 2009) that pairs similar (in terms of size and/or related covariates) clusters and randomly assigns one cluster from each pair to treatment. In our setting, it is a special case of the IBR experiments where each block contains two clusters and the marginal assignment probability $q = \frac{1}{2}$. The correlation matrix of the assignments is a block diagonal matrix with each block being a 2×2 dimension, the diagonal entries being 1, and the off-diagonal entries being -1. By Corollary 4.6 with $k = 2$, the approximation ratio of the pair experiment is $5/2$, which is slightly larger than the $7/3$ approximation ratio guarantee for the optimal IBR experiment. On the other hand, for the special case of equal cluster sizes (and assuming n is even), the worst-case variance of the pair experiment is $\frac{n}{2}$. Thus, Corollary 3.2 implies that the multiplicative gap is 2 relative of the optimal IBR experiment. The pair experiment is not asymptotically optimal in this case as n grows large, whereas the optimal IBR experiment is asymptotically optimal under mild regularity conditions on cluster sizes, as we discuss in Section 4.2. Finally, for the special case of $q = \frac{1}{m}$ with some integer m, we can consider a generation of the pair-matching experiment that groups m clusters of similar size together and randomly assigns one cluster from each group to treatment. As we discussed earlier, this is equivalent to the m-partition experiment, and Corollary 4.4 provides approximation ratio guarantees for such an experiment provided that m is small (e.g., $m \leq 10$).

Finally, in Lemma 4.7, we show that the $f_q(k)$ function grows linearly in k; we will use this rate property in our asymptotic analysis in Section 4.2 shortly. We prove Lemma 4.7 in Appendix B.7.

Lemma 4.7. $f_q(k) \leq \frac{1}{4}(k + C(q))$*, where* $C(q)$ *is a constant that depends only on* q*.*

4.2 Asymptotic Optimality

In our asymptotic analysis, we consider the regime where the number of clusters n grows to infinity and there is no dominating cluster, in the sense that $w_1^2 = o\big(\sum_{i\in[n]} w_i^2\big)$. Since $V^{\text{OPT}} = \Theta\big(\sum_{i\in[n]} w_i^2\big)$ by Lemma 4.2, this intuitively means that no cluster is large enough to have a substantial effect on the variance of the estimator.

Under this regime, Lemma 4.3 immediately implies that any k-partition IBR experiment with $k = \Theta\left(\sqrt{\frac{\sum_{i\in[n]} w_i^2}{w_1^2}}\right)$ is asymptotically optimal. To see this, note that by Lemmas 4.2, 4.3, and 4.7, we have

$$
\begin{aligned}
\frac{V^{\mathrm{DP}} - V^{\mathrm{LB}}}{V^{\mathrm{LB}}} \le \frac{V^{k} - V^{\mathrm{LB}}}{V^{\mathrm{LB}}} &\le \big(k + C(q)\big)\left(\frac{1}{k} + \frac{k-1}{k}\frac{w_1^2}{\sum_{i\in[n]} w_i^2}\right) - 1 \\
&\le \frac{C(q)}{k} + k \cdot \frac{w_1^2}{\sum_{i\in[n]} w_i^2} + C(q) \cdot \frac{w_1^2}{\sum_{i\in[n]} w_i^2} \\
&= O\left(\sqrt{\frac{w_1^2}{\sum_{i\in[n]} w_i^2}}\right) \to 0,
\end{aligned}
$$

where the last equality is attained, e.g., if we let k be the integer closest to $\sqrt{\frac{\sum_{i\in[n]} w_i^2}{w_1^2}}$. Since the optimal IBR experiment achieves a smaller variance than the k-partition IBR experiment, it is asymptotically optimal as well, as we establish next.

Theorem 4.8 (*Asymptotic Performance of the Optimal IBR Experiment*)**.** *The optimal IBR experiment is asymptotically optimal when the number of clusters n grows large and $w_1^2 = o\big(\sum_{i\in[n]} w_i^2\big)$. Moreover, the convergence rate satisfies*

$$
\frac{V^{\mathrm{DP}} - V^{\mathrm{LB}}}{V^{\mathrm{LB}}} = O\left(\sqrt{\frac{w_1^2}{\sum_{i\in[n]} w_i^2}}\right) \to 0.
$$

We next illustrate by Example 4.1 that the aforementioned "no dominating cluster" condition is also necessary for any IBR experiment to be asymptotically optimal.

Example 4.1. Consider a problem instance where the marginal assignment probability $q = \frac{1}{2}$ and the cluster sizes form a geometric sequence, i.e., $w_i = \beta^{n-i}$ for cluster $i \in [n]$, and we let $\beta = \frac{5}{4}$. It can be shown that in the optimal partition, all blocks contain four clusters. More precisely, suppose that n is divisible by four; then every block h contains clusters $4h-3$ to $4h$ for $1 \le h \le n/4$. Note that each block essentially contains the same clusters up to a scaling. Since the optimal IBR experiment randomly assigns half of the clusters in a block to treatment, it increases the worst-case variance of each block by a constant fraction of 10.7% compared to an experiment that assigns treatment to clusters in a block in an optimal way, and does so independently across blocks. Thus, the optimal IBR experiment cannot be asymptotically optimal as n grows large. The no dominating cluster condition is violated because $\sum_{i=1}^{n} w_i^2 = \frac{\beta^{2n}-1}{\beta^2-1}$ and $w_1^2 = \beta^{2n-2}$ have the same order. We provide further details in Appendix B.8.

4.2.1 Asymptotic Optimality in the High-Multiplicity Model

The asymptotic optimality in Theorem 4.8 is for a general setting, and we can obtain a stronger result under additional assumptions. Specifically, for a high-multiplicity model where cluster sizes take values in a finite set and the number of clusters of each size is a fixed proportion of the total number of clusters n, a simple IBR experiment that has one block for each cluster size increases the worst-case variance only by a constant that is independent of n, as we show in Theorem 4.9.

Theorem 4.9 (*Improved Performance Guarantee in the High-Multiplicity Model*)**.** *Suppose that cluster sizes take only K finite values $(w_i)_{i\in[K]}$ with $w_1 \geq w_2 \geq \cdots \geq w_K$. Let n_k be the number of clusters of size w_k and consider a simple IBR experiment that has one block for each cluster size. The worst-case variance V of this experiment satisfies*

$$V - V^{\mathrm{LB}} \leq \frac{1}{4}\sum_{k\in[K]} w_k^2 \cdot \left(\min\left\{n_k, \frac{w_1^2 n}{w_k^2 n_k}\right\} + \frac{2}{q(1-q)} + 4\right),$$

where the right-hand side scales with the total number of clusters n at most at a square-root rate (whereas, the lower bound satisfies $V^{\mathrm{LB}} = \Theta(n)$).

If, in addition, each $n_k = \alpha_k n$ is a fixed proportion $\alpha_k \in (0,1)$ of the total number of clusters n, then

$$V - V^{\mathrm{LB}} \leq \sum_{k\in[K]} \left\{\frac{w_1^2}{4\alpha_k} + w_k^2\left(\frac{1}{2q(1-q)} + 1\right)\right\},$$

which is a constant independent of n.

We prove Theorem 4.9 in Appendix B.9. When each block contains clusters of equal size, there is no loss of optimality from ignoring the cluster size differences within a block (i.e., the second source of the performance loss). Thus, Theorem 4.9 indicates that the loss from independent assignments across blocks (i.e., the first source of the performance loss) can be made small (and in fact asymptotically negligible) with an IBR experiment. Intuitively, if an experiment is close to the optimal experiment, the worst-case potential outcomes of the two experiments are close as well. Note that we have a complete characterization of the worst-case potential outcome with an IBR experiment (Proposition 3.1 and Lemma 3.3). In our proof, we bound the worst-case variance gap between the simple IBR experiment and the optimal experiment by considering both experiments against the worst-case potential outcome of the simple IBR experiment, and by using the fact that the correlation matrix of any experiment is positive semidefinite.

4.2.2 Simplicity versus Complexity: The IBR Experiment with a Logarithmic Partition

Although the k-partition IBR experiment with $k = \Theta\Big(\sqrt{\sum_{i\in[n]} w_i^2/w_1^2}\Big)$ is asymptotically optimal under the no dominating cluster condition, to compute its partition, we need to solve a DP for every group that contains $k = \Omega(1)$ clusters. We next show that a very simple logarithmic partition that requires no explicit optimization is sufficient for an IBR experiment to be asymptotically optimal. However, the convergence rate of the corresponding experiment could be slower and we require a slightly stronger version of the no dominating cluster condition.

Let us start by introducing the slightly stronger version of the no dominating cluster condition needed for our analysis of the aforementioned experiment: $\frac{\sum_{i=1}^n w_i^2}{w_1^2} = \Omega\big(n^c\big)$ for some constant $c > 0$. We consider a simple logarithmic partition in which the ratio of the largest to the smallest cluster sizes in a block is upper-bounded across blocks. In particular, we define two parameters $\delta_1, \delta_2 \in (0, 1)$ to be specified later. We first include all clusters with sizes $w_i \le \bar{w} \triangleq \sqrt{\frac{\sum_{i=1}^n w_i^2}{n^{1+\delta_1}}}$ in one block, and label this as block zero. The clusters in this block are small enough to have little effect on the variance of the estimator. We then iteratively go through the remaining clusters in decreasing order of size to create blocks. Specifically, in each step we focus on the clusters that are not yet assigned to a block, pick the largest one, and include all of the clusters whose sizes are at least $\frac{1}{\alpha}$ times the size of this cluster in a block. Here, α is a parameter given by $\alpha = 1 + n^{-\delta_2}$. We label these blocks from one to K, in decreasing order of size from the largest cluster in each block. Theorem 4.10 shows that such a logarithmic partition induces an asymptotically optimal IBR experiment when δ_1 and δ_2 are chosen properly.

Theorem 4.10 (*Asymptotic Performance of the IBR Experiment with a Logarithmic Partition*)**.** *Suppose that $\sum_{i=1}^n w_i^2/w_1^2 = \Omega(n^c)$ with some constant $c > 0$. Consider the above logarithmic partition with parameters $\delta_1 = \delta_2 = c/4$, and let V denote the worst-case variance of the logarithmic-partitioning IBR experiment. This logarithmic-partitioning IBR experiment is asymptotically optimal, and the convergence rate satisfies*

$$\frac{V - V^{\mathrm{LB}}}{V^{\mathrm{LB}}} = O\left(n^{-\frac{c}{4}} \ln n\right).$$

We prove Theorem 4.10 in Appendix B.10. Because of the specific partition we chose, clusters in the same block have similar sizes (the ratio of sizes of the largest to the smallest clusters in a block is at most α, which goes to one as n grows large). This makes the second source of performance loss small. In our proof, we consider a perturbed problem where we decrease all cluster sizes in a block to the minimum cluster size of the block. This is only a small perturbation and does not change the worst-case variance of an experiment much, as clusters in a block are similar in size. After the perturbation, clusters in a block have equal sizes, and we can adopt the analysis of Theorem 4.9.

5 Numerical Examples

In this section, we examine the performance of our IBR experiment on two numerical examples: a synthetic example with randomly generated instances (Section 5.1) and a data-driven example based on real Airbnb data (Section 5.2). Moreover, we compare its performance with heuristic experiments such as (i) independent cluster-based randomization, (ii) the experiment that uniformly at random assigns a fixed fraction of the clusters to treatment, and (iii) the (generalized) pair-matching experiment. We illustrate that the worst-case variance of our IBR experiment is small compared to these heuristic experiments and is close to the worst-case variance of the optimal cluster-based experiment.

In Appendix E, we consider a third example based on Facebook data. In addition to comparing the worst-case variances, we assume that potential outcomes are random draws from a given distribution, and numerically compare the "average case" performances of these experiments. Our results indicate that the IBR experiment still reduces the variance substantially relative to independent cluster-based randomization and improves upon other heuristic experiments.

5.1 Synthetic Examples with a Small Number of Clusters

We first consider randomly generated instances of the cluster-based experimental design problem with number of clusters $n \in \{6, 8, 10, 12\}$ and marginal assignment probability $q \in \{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}\}$. For each fixed value of n and q, we generate *i.i.d.* cluster sizes, each uniformly at random between 1 and 100, i.e., $w_i \sim \text{Unif}\{1, \ldots, 100\}$. We create 10^5 samples for Monte Carlo simulation. For each sample, we calculate: (a) the worst-case variance V^{DP} of the optimal IBR experiment (by solving the DP of Section 3), (b) the worst-case variance V^{OPT} of the optimal cluster-based experiment, and (c) the worst-case variance $V^{k=k^*}$ of the k-partition IBR experiment using the optimal value of k (which we denote by k^*) that attains the minimum approximation ratio bound in Theorem 4.5. As discussed earlier, the optimization problem of solving the optimal worst-case variance V^{OPT} (and the corresponding experiment) can be reformulated as a linear program with exponentially many variables/constraints in (6) in Appendix B.3; the aforementioned values of n are small enough so that we are able to solve this linear program. The k-partition IBR experiments are our main analytical tool in Section 4. Recall that from Table 1, the optimal value k^* is equal to 4 for $q = \frac{1}{2}$, and equal to $\frac{1}{q}$ for the other values of q. Finally, for the case of q equal to $\frac{1}{2}$, we also calculate: (d) the worst-case variance $V^{k=2}$ of the k-partition IBR experiment with $k = \frac{1}{q} = 2$ (which corresponds to the pair-matching experiment in Remark 4.4).

In Figure 3, we draw the box plots of the ratios $\frac{V^{\text{DP}}}{V^{\text{OPT}}}$ and $\frac{V^{k=k^*}}{V^{\text{OPT}}}$ over the 10^5 samples for each fixed q and n, and we report the max values of these ratios in Table 2. Overall, for each fixed q, the optimal IBR experiment performs quite well and in most instances it increases the worst-case variance only by at most 50%. Notably, this is substantially better than the approximation ratio guarantee in Theorem 4.5 (and Table 1).

q	1/2	1/3	1/4	1/5
max of $V^{\text{DP}}/V^{\text{OPT}}$	1.521	1.491	1.506	1.470
max of $V^{k=k^*}/V^{\text{OPT}}$	1.617	1.630	1.549	1.527

q	1/2
max of $V^{k=2}/V^{\text{OPT}}$	2.065

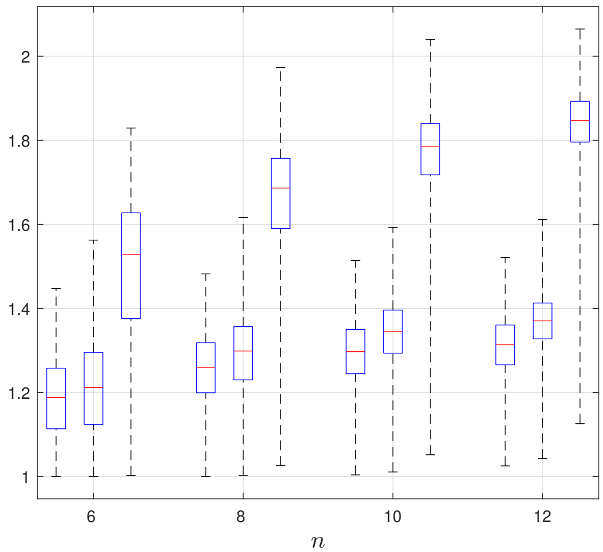
Table 2: Maximum values of $V^{\text{DP}}/V^{\text{OPT}}$ and V^k/V^{OPT} for randomly generated examples for different q's.

5.2 The Airbnb Example

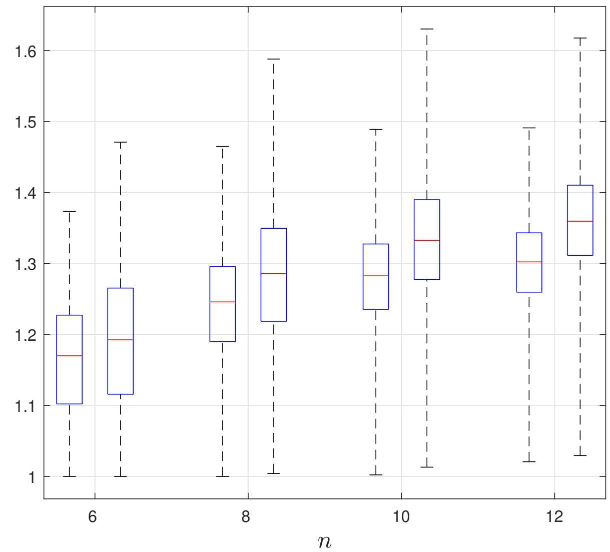
Next, we examine a data-driven example based on Airbnb data. While our Airbnb example is relatively smaller, we also consider another example that involves a larger Facebook subnetwork of one hundred US universities in Appendix E. The Airbnb example uses data from the Inside Airbnb website (InsideAirbnb 2016). This website collects detailed information of listings from the online hospitality platform Airbnb, including the longitude/latitude coordinates of the listing, the room type, capacity, price-per-night, and the minimum and maximum number of nights to stay. We use all the listings in the Bay area (California, United States) that are collected by Inside Airbnb in our experiment. There are in total $16,010$ listings, which include listings in San Francisco, Oakland, San Mateo County, and Santa Clara County. Figure 5 in Appendix D depicts the geographical locations of these listings.

Extracting Clusters We use a similar approach as in Holtz and Aral (2020) to partition the listings into well-separated clusters. To do so, we first construct an interference network among the listings, where each node represents a listing and there is an edge between two listings if they are likely to substitute (and hence interfere with) each other. Specifically, we assume that there is an edge between two listings if *all* of the following six criteria are satisfied:

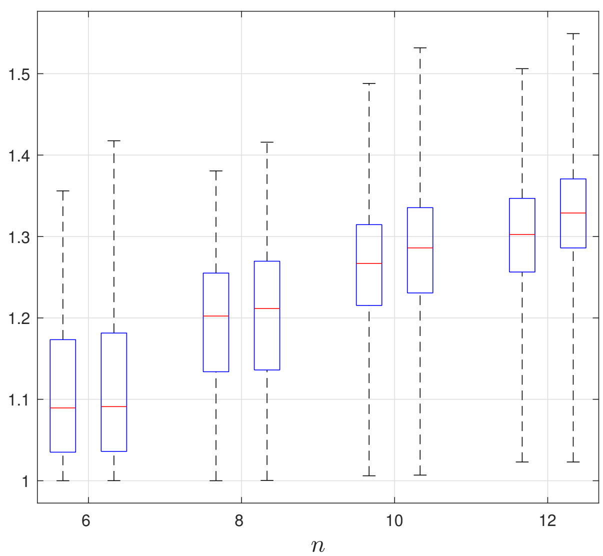
1. The two listings are either both in San Francisco or both outside San Francisco;
2. The two listings are within 1 mile of straight-line distance from each other if both are in San Francisco, and are within 5 miles of straight-line distance if both are outside San Francisco;



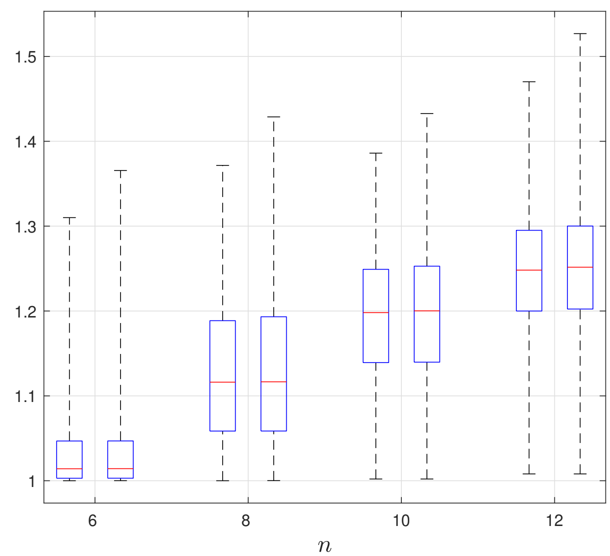
(a) $q = \frac{1}{2}$



(b) $q = \frac{1}{3}$



(c) $q = \frac{1}{4}$



(d) $q = \frac{1}{5}$

Figure 3: For each fixed number of clusters n: (a) draws the box plots of the ratios $V^{\text{DP}}/V^{\text{OPT}}$ (left), $V^{k=4}/V^{\text{OPT}}$ (middle) and $V^{k=2}/V^{\text{OPT}}$ (right) with $q = \frac{1}{2}$, and (b)–(d) draw the box plots of the ratios $V^{\text{DP}}/V^{\text{OPT}}$ (left) and $V^{k=m}/V^{\text{OPT}}$ with $m = \frac{1}{q}$ (right) with $q = \frac{1}{3}$, $\frac{1}{4}$ and $\frac{1}{5}$, respectively. In each box plot, the central red edge indicates the median, the bottom and top blue edges of the box indicate the 25th and 75th percentiles, and the bottom and top black edges outside the box indicate the minimum and maximum extreme values of the ratio, all over the 10^5 samples.

3. The two listings have the same room type;
4. The two listings have overlap with respect to the feasible number of nights to stay (i.e., for each listing, its minimum number of nights to stay is no larger than the maximum number of nights to stay of the other listing);

5. The capacity of one listing is no more than twice the capacity of the other listing;
6. The price-per-night of one listing is no higher than twice the price-per-night of the other one.

In other words, we assume that San Francisco is a disjoint submarket. Moreover, two listings interfere with each other if they are geographically close, have the same room type, have overlap in the number of nights to stay, and are comparable in terms of capacity and price-per-night.

We then apply the well-known Louvain algorithm (Blondel et al. 2008) to partition the interference network into clusters of varying sizes.[13] We choose the ten largest clusters from the output for our experiment; the sizes (i.e., number of listings) of these ten clusters are:

$$2566, 2100, 2093, 1908, 1629, 1535, 1390, 1181, 590, 518.$$

Note that these ten clusters already cover 97% of all the listings, and none of the remaining clusters is larger than 4.8% of the largest cluster. We visualize these ten clusters in Figure 6 and provide more details of these clusters in Table 4 in Appendix D.

The $n = 10$ clusters separate well from each other. Specifically, 77% of the listings connect only to other listings in the same cluster. In most of the remaining listings, only a small fraction of their connections are from a different cluster; please refer to Figure 7 in Appendix D for a histogram of fraction of connections from a different cluster for these listings. In what follows, we assume that the interference among clusters is relatively small and can be ignored, and we evaluate the worst-case variances of various experiments on these ten clusters. This assumption is particularly valid when we consider exposure models where the outcome of a listing can be affected by its neighbors only if at least a certain fraction of them have a different treatment/control assignment.

Comparing Cluster-Based Experiments We consider the case where the marginal assignment probability q is equal to $\frac{1}{2}$, and we assume that the upper bounds w_{i1} and w_{i0} of the cluster-level treatment and control potential outcomes (and hence the upper bound w_i in (1)) are both proportional to the number of listings in cluster i.

Although we can solve an optimal cluster-based experiment with $n = 10$ clusters, this experiment results in a complex randomized assignment (see Table 5 in Appendix D). Specifically, the experiment randomizes over 53 different possible assignment vectors (where the number of treated

[13]This is a classic approach to extracting clusters in large networks. It tries to construct a partition of nodes by maximizing modularity, i.e., the fraction of edges that remain within cluster/partition relative to a random distribution of edges. Modularity maximization is itself a computationally challenging problem; thus, the algorithm relies on a greedy heuristic for this purpose. See Blondel et al. (2008) for details.

clusters varies between 4 and 6), and chooses different probabilities for these vectors without following any clear patterns. It even deliberately introduces some amount of positive correlation between small clusters to attain a larger negative correlation between some pairs of large and small clusters. In particular, the correlation matrix of this assignment is:

$$\Sigma^{\mathrm{OPT}} = \begin{pmatrix} 1 & -0.251 & -0.250 & -0.220 & -0.186 & -0.172 & -0.155 & -0.145 & -0.068 & -0.073 \\ -0.251 & 1 & -0.172 & -0.162 & -0.134 & -0.126 & -0.109 & -0.066 & -0.031 & -0.044 \\ -0.250 & -0.172 & 1 & -0.161 & -0.133 & -0.126 & -0.108 & -0.131 & -0.057 & -0.045 \\ -0.220 & -0.162 & -0.161 & 1 & -0.112 & -0.102 & -0.099 & -0.097 & -0.034 & -0.049 \\ -0.186 & -0.134 & -0.133 & -0.112 & 1 & -0.086 & -0.067 & -0.060 & -0.034 & -0.057 \\ -0.172 & -0.126 & -0.126 & -0.102 & -0.086 & 1 & -0.071 & -0.065 & -0.036 & 0.007 \\ -0.155 & -0.109 & -0.108 & -0.099 & -0.067 & -0.071 & 1 & -0.058 & -0.040 & 0.008 \\ -0.145 & -0.066 & -0.131 & -0.097 & -0.060 & -0.065 & -0.058 & 1 & -0.047 & 0.009 \\ -0.068 & -0.031 & -0.057 & -0.034 & -0.034 & -0.036 & -0.040 & -0.047 & 1 & 0.004 \\ -0.073 & -0.044 & -0.045 & -0.049 & -0.057 & 0.007 & 0.008 & 0.009 & 0.004 & 1 \end{pmatrix}.$$

As can be seen from the above matrix and the details in Appendix D.2, the optimal experiment has a fairly complicated correlation structure that makes it hard to interpret.

By contrast, the optimal IBR experiment—which is obtained by solving the DP presented earlier—has only three blocks; it simply places the first four clusters in one block, the next four clusters in a second block, and the last two clusters in the final block. Then, it randomly assigns half of the clusters (chosen uniformly at random) in each block to treatment. Thus, unlike the optimal cluster-based experiment, the optimal IBR experiment is easy to interpret: it pools clusters of similar size together in the same block, and then in each block pretends that the cluster sizes are the same and runs the optimal cluster-based experiment accordingly inside the block. The worst-case variance of this experiment increases the worst-case variance by $\frac{V^{\mathrm{DP}}-V^{\mathrm{OPT}}}{V^{\mathrm{OPT}}} = 30.8\%$ relative to the optimal cluster-based experiment.

We also compare our design with three natural heuristics:

1. `HALF`: The simple experiment that randomly assigns half of the clusters to treatment (i.e., an IBR experiment with only one block);
2. `PAIR`: The pair-matching experiment that sorts the clusters based on their sizes, pairs each cluster with an odd index in the sorted list to the next cluster with an even index, and finally randomly assigns one cluster from each pair to treatment (i.e., an IBR experiment with each block containing two clusters);
3. `IND`: The naive experiment with an independent cluster-based assignment (which assigns each

cluster to treatment or control independently with probability 1/2).

The three experiments increase the worst-case variance by $\frac{V^{\text{half}}-V^{\text{OPT}}}{V^{\text{OPT}}} = 50.8\%$, $\frac{V^{\text{pair}}-V^{\text{OPT}}}{V^{\text{OPT}}} = 87.5\%$, and $\frac{V^{\text{ind}}-V^{\text{OPT}}}{V^{\text{OPT}}} = 229.9\%$, respectively. This example illustrates that not only is the IBR experiment considerably easier to compute and implement than the optimal cluster-based assignment, but it also admits a worst-case variance that increases that of the optimal experiment by only a mild amount. At the same time, other simple heuristic designs yield considerably higher worst-case variances. For further details on this example see Appendix D.

6 Conclusion and Further Directions

We have considered the problem of designing a randomized experiment over a set of disjoint clusters to minimize the variance of an unbiased Horvitz–Thompson estimator that estimates the total market effect. We formulate the problem as robust optimization against the adversarial selection of potential outcomes. An optimal cluster-based assignment is computationally expensive to solve, and can be difficult to implement due to the required complicated correlation structure. Motivated by this, we develop a family of simple independent block randomization experiments that are easy to compute and interpret. These experiments are optimal when all clusters have identical sizes. More generally, we show that IBR experiments are asymptotically optimal (in the number of clusters) under a mild no dominating cluster condition and constitute a good approximation for any problem instance when the marginal assignment probability is not very tiny. In general settings, the suboptimality originates from the loss both due to independence across blocks and due to ignoring the cluster size differences within a block. Our results indicate that this suboptimality can be made small with a careful partitioning of clusters into blocks.

In our model, we assume that the marginal assignment probability q is given, whereas, in practice, the decision maker may also want to incorporate q as a decision variable. The corresponding optimal cluster-based experiment becomes computationally even harder to solve because it no longer admits a linear programming formulation, and the objective is generally not a convex function of the marginal probability q. Since it is fairly easy to solve an optimal IBR experiment with a given q (by solving the DP in (4)), we can do 1-dimensional search to find an optimal value for q such that the worst-case variance of the optimal IBR experiment is minimized. Our analysis implies that such an IBR experiment is approximately and asymptotically optimal even compared to the optimal cluster-based experiment that optimizes over the marginal probability q as well.

Acknowledgements

We thank Gabriel Weintraub, the area editor Itai Ashlagi, and anonymous associate editor and referees for providing a number of valuable comments that greatly improved the paper.

References

Aronow, P. M. and Middleton, J. A. (2013), 'A class of unbiased estimators of the average treatment effect in randomized experiments', *Journal of Causal Inference* **1**(1), 135–154.

Aronow, P. M. and Samii, C. (2017), 'Estimating average causal effects under general interference, with application to a social network experiment', *The Annals of Applied Statistics* **11**(4), 1912–1947.

Asmussen, S. and Glynn, P. W. (2007), *Stochastic simulation: algorithms and analysis*, Vol. 57, Springer Science & Business Media.

Bai, Y. (2022), 'Optimality of matched-pair designs in randomized controlled trials', *Available at SSRN 3483834* .

Barajas, J., Akella, R., Holtan, M. and Flores, A. (2016), 'Experimental designs and estimation for online display advertising attribution in marketplaces', *Marketing Science* **35**(3), 465–483.

Basse, G. W., Soufiani, H. A. and Lambert, D. (2016), Randomization and the pernicious effects of limited budgets on auction experiments, *in* 'Artificial Intelligence and Statistics', PMLR, pp. 1412–1420.

Berger, J. O. (2013), *Statistical decision theory and Bayesian analysis*, Springer Science & Business Media.

Bertsimas, D., Johnson, M. and Kallus, N. (2015), 'The power of optimization over randomization in designing experiments involving small samples', *Operations Research* **63**(4), 868–876.

Bertsimas, D., Korolko, N. and Weinstein, A. M. (2019), 'Covariate-adaptive optimization in online clinical trials', *Operations Research* **67**(4), 1150–1161.

Bhat, N., Farias, V. F., Moallemi, C. C. and Sinha, D. (2020), 'Near-optimal ab testing', *Management Science* **66**(10), 4477–4495.

Bibaut, A., Chambaz, A., Dimakopoulou, M., Kallus, N. and van der Laan, M. (2021), 'Post-contextual-bandit inference', *arXiv preprint arXiv:2106.00418* .

Blondel, V. D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. (2008), 'Fast unfolding of communities in large networks', *Journal of statistical mechanics: theory and experiment* **2008**(10), P10008.

Bojinov, I., Simchi-Levi, D. and Zhao, J. (2020), 'Design and analysis of switchback experiments', *Available at SSRN 3684168* .

Bright, I., Delarue, A. and Lobel, I. (2022), 'Reducing marketplace interference bias via shadow prices', *arXiv preprint arXiv:2205.02274* .

Chamandy, N. (2016), 'Lyft engineering: Experimentation in a ridesharing marketplace'. Available at https://eng.lyft.com/experimentation-in-a-ridesharing-marketplace-b39db027a66e#.djox1933t.

Cui, R., Li, J. and Zhang, D. J. (2020), 'Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb', *Management Science* **66**(3), 1071–1094.

Eckles, D., Karrer, B. and Ugander, J. (2017), 'Design and analysis of experiments in networks: Reducing bias from interference', *Journal of Causal Inference* **5**(1).

Fisher, R. A. (1935), *The design of experiments*, Oliver & Boyd.

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S. and Athey, S. (2019), 'Confidence intervals for policy evaluation in adaptive experiments', *arXiv preprint arXiv:1911.02768* .

Harshaw, C., Sävje, F., Spielman, D. and Zhang, P. (2019), 'Balancing covariates in randomized experiments using the gram-schmidt walk', *arXiv preprint arXiv:1911.03071* .

Higgins, M. J., Sävje, F. and Sekhon, J. S. (2016), 'Improving massive experiments with threshold blocking', *Proceedings of the National Academy of Sciences* **113**(27), 7369–7376.

Holtz, D. and Aral, S. (2020), 'Limiting bias from test-control interference in online marketplace experiments', *arXiv preprint arXiv:2004.12162* .

Holtz, D., Lobel, R., Liskovich, I. and Aral, S. (2020), 'Reducing interference bias in online marketplace pricing experiments', *arXiv preprint arXiv:2004.12489* .

Horvitz, D. G. and Thompson, D. J. (1952), 'A generalization of sampling without replacement from a finite universe', *Journal of the American statistical Association* **47**(260), 663–685.

Hudgens, M. G. and Halloran, M. E. (2008), 'Toward causal inference with interference', *Journal of the American Statistical Association* **103**(482), 832–842.

Imai, K., King, G. and Nall, C. (2009), 'The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation', *Statistical Science* **24**(1), 29–53.

Imbens, G. W. and Rubin, D. B. (2015), *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.

InsideAirbnb (2016). Available at http://insideairbnb.com/. Accessed: 2022-04-26.

Jin, Y. and Ba, S. (2021), 'Towards optimal variance reduction in online controlled experiments', *arXiv preprint arXiv:2110.13406* .

Johari, R., Li, H., Liskovich, I. and Weintraub, G. (2020), 'Experimental design in two-sided platforms: An analysis of bias', *arXiv preprint arXiv:2002.05670* .

Kallus, N. (2018), 'Optimal a priori balance in the design of controlled experiments', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**(1), 85–112.

Karrer, B., Shi, L., Bhole, M., Goldman, M., Palmer, T., Gelman, C., Konutgan, M. and Sun, F. (2020), 'Network experimentation at scale', *arXiv preprint arXiv:2012.08591* .

Kohavi, R., Tang, D. and Xu, Y. (2020), *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, Cambridge University Press.

Koutra, V. (2017), Designing experiments on networks, PhD thesis, University of Southampton.

Lattimore, T. and Szepesvári, C. (2020), *Bandit algorithms*, Cambridge University Press.

Li, S., Johari, R., Wager, S. and Xu, K. (2023), 'Experimenting under stochastic congestion', *arXiv preprint arXiv:2302.12093* .

Lin, W. (2013), 'Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique', *The Annals of Applied Statistics* **7**(1), 295–318.

Owen, A. B. (2020), 'A first course in experimental design notes from stat 263/363', *Lecture notes* . Available at https://statweb.stanford.edu/~owen/courses/363/doenotes.pdf.

Pouget-Abadie, J., Aydin, K., Schudy, W., Brodersen, K. and Mirrokni, V. (2019), 'Variance reduction in bipartite experiments through correlation clustering', *Advances in Neural Information Processing Systems* **32**, 13309–13319.

Rachev, S. T. and Rüschendorf, L. (1998), *Mass Transportation Problems: Volume I: Theory*, Vol. 1, Springer Science & Business Media.

Rossi, R. A. and Ahmed, N. K. (2015), 'The network data repository'. Available at http://networkrepository.com/socfb.php.

Rüschendorf, L. and Uckelmann, L. (2002), Variance minimization and random variables with constant sum, *in* 'Distributions with given marginals and statistical modelling', Springer, pp. 211–222.

Slivkins, A. (2019), 'Introduction to multi-armed bandits', *Foundations and Trends® in Machine Learning* **12**(1-2), 1–286.

Spielman, D. A. and Teng, S.-H. (2004), Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems, *in* 'Proceedings of the thirty-sixth annual ACM symposium on Theory of computing', pp. 81–90.

Traud, A. L., Mucha, P. J. and Porter, M. A. (2012), 'Social structure of facebook networks', *Physica A: Statistical Mechanics and its Applications* **391**(16), 4165–4180.

Ugander, J., Karrer, B., Backstrom, L. and Kleinberg, J. (2013), Graph cluster randomization: Network exposure to multiple universes, *in* 'Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining', pp. 329–337.

Ugander, J. and Yin, H. (2020), 'Randomized graph cluster randomization', *arXiv preprint arXiv:2009.02297* .

Wager, S. and Xu, K. (2021), 'Experimenting in equilibrium', *Management Science* .

Williamson, D. P. and Shmoys, D. B. (2011), *The design of approximation algorithms*, Cambridge university press.

Zigler, C. M. and Papadogeorgou, G. (2021), 'Bipartite causal inference with interference', *Statistical science: a review journal of the Institute of Mathematical Statistics* **36**(1), 109.

A More on Related Literature

Foundations of Experimental Design Experimental design has found far-reaching applications to guide decision making in areas ranging from medical trials to social sciences, and recently in online marketplaces and social networks. It is grounded in causal inference, but instead of inferring causality from purely observational data, a decision maker can choose how to gather data to gain more statistical efficiency and more convincing empirical evidence of causality. The process often involves randomization, and a more careful design of randomization based on optimization helps to maximize the statistical power. Great expositions of related topics include Owen (2020), Kohavi et al. (2020), and Imbens and Rubin (2015).

Regression Adjustment in Experimental Design There is a rich literature on variance reduction via regression adjustment with covariate information in randomized experiments (e.g., Lin 2013, Jin and Ba 2021, and the references therein). These works fix the joint treatment assignment to be completely random and consider the best analysis (i.e., the estimator) to minimize the variance in the asymptotic regime (under certain settings). They show that agnostic regression can reduce variance. Our work takes the opposite direction. Specifically, we fix the estimator (i.e., we deliberately use a common estimator—the simple unadjusted Horvitz-Thompson estimator) and study the best design. We view this as a building block that might eventually enable combining the two steps and optimizing them jointly. Another distinction is that we provide a bound on performance loss for any problem instance, and such a bound implies that our design is asymptotically optimal under the adversarial model. By contrast, the regression adjustment literature has mainly focused on the asymptotic analysis.

Pair-Matching Experiments The pair-matching experiment pairs similar (in terms of size and/or related covariates) clusters and randomly assigns one cluster from each pair to treatment. It is a special case of the IBR experiments where each block contains two clusters and the marginal assignment probability $q = \frac{1}{2}$. Imai et al. (2009) recommend a pair-matching experiment if cluster-based randomization is used. While the authors do not provide a theoretical justification for this recommendation, they offer empirical evidence for the usefulness of this design based on comparison with other heuristic experiments. Our work, by contrast, provides a theoretical foundation for the aforementioned design. Specifically, we show that an optimal IBR experiment can achieve much of the benefit from the optimal (correlated) randomized assignment across clusters, and is asymptotically optimal when the number of clusters grows large under a mild regularity assumption on cluster sizes. On the other hand, although the pair-matching experiment is not asymptotically optimal, its worst-case variance is guaranteed to be within $\frac{5}{2}$ of the variance of the optimal experiment (Remark 4.4), which is slightly larger than that of the optimal IBR experiment. Finally, Bai (2022) shows that certain pair-matching experiments are optimal under stratified randomization (our IBR experiments take the same form; please refer to Remark 3.1) using a different sampling-based model where units' potential outcomes are independently sampled.

Alternative Approaches to Experimental Design Other recent papers have considered models in which a covariate of a unit is correlated with the potential outcome in a certain way. The optimal experiment usually involves covariate balancing, such that the treatment and control groups are similar in terms of the covariates (Bertsimas et al. 2015, Bertsimas et al. 2019, Kallus 2018, Bhat et al. 2020, and Harshaw et al. 2019). We instead work with a model of potential outcomes with minimal assumptions on their ranges (nevertheless, these ranges can still be inferred from the covariate information).

The rich literature on online learning and multi-armed bandits (see, e.g., Lattimore and Szepesvári 2020 and Slivkins 2019 for surveys) can also be viewed as a form of adaptive and sequential experimental design. There, the decision maker is allowed to switch between variants (i.e., arms), the system is assumed to be stationary and have rapid feedback (i.e., no carryover effect), and the objective is to find the best variant with minimum cumulative regret or number of trials; see, e.g., Hadad et al. (2019) and Bibaut et al. (2021). Finally, we point out that the experimental design problem of minimizing the variance of an (unbiased) estimator is also very related to various variance-reduction methods in the simulation literature (e.g., Asmussen and Glynn 2007).

B Proofs

B.1 Details of Remark 2.2

Suppose there are n Bernoulli random variables $Z = (Z_i)_{i=1}^n$, each with a marginal probability $\mathbb{P}[Z_i] = \frac{1}{2}$. The joint distribution of Z that minimizes the variance $\mathbb{V}\mathrm{ar}\Big[\sum_{i=1}^n w_i Z_i\Big]$ (with known weights w_i) is as follows: partition the weights $\{w_i\}$ into two groups as even as possible, then randomly select one group with probability $\frac{1}{2}$, let $Z_i = 1$ for all units in that group and $Z_i = 0$ for the rest units.

Proof. For any joint distribution of $Z = (Z_i)_{i\in[n]}$ with $\mathbb{P}[Z_i] = \frac{1}{2}$ for all $i \in [n]$, the mean value is fixed, i.e., $\mathbb{E}[\sum_{i\in[n]} w_i Z_i] = \frac{1}{2}\sum_{i\in[n]} w_i$. For a set $S \subseteq [n]$, let $w(S) \triangleq \sum_{i\in S} w_i$ denote the sum of weights of units in set S and $d(S) \triangleq \left| w(S) - \frac{1}{2}\sum_{i\in[n]} w_i \right|$ denote the absolute difference between the weight of set S and the mean value $\frac{1}{2}\sum_{i\in[n]} w_i$. Let $S^* \triangleq \operatorname{argmin}_{S\subseteq[n]} d(S)$ and $d^* = d(S^*)$; clearly, we have $w(S^*) = w\big([n]\backslash S^*\big) \le w(S)$ for any set $S \subseteq [n]$, and $\big(S^*, [n]\backslash S^*\big)$ is a balanced partition of $\{w_i\}$.

On the other hand, the variance satisfies

$$\mathbb{V}\mathrm{ar}\Big[\sum_{i=1}^{n} w_i Z_i\Big] = \mathbb{E}\left[d\big(\{i : Z_i = 1\}\big)^2\right] \ge d^*,$$

where the inequality holds by the definition of d^*. Furthermore, if all units $i \in S^*$ take $Z_i = 1$ and the rest take $Z_i = 0$ with probability $\frac{1}{2}$, and vice versa with probability $\frac{1}{2}$, the equality is attained; thus, this joint random assignment satisfies $\mathbb{P}[Z_i] = \frac{1}{2}$ for all units $i \in [n]$, and it minimizes the variance. $\square$

B.2 Details of Remark 2.3

Suppose that the uncertainty set of each cluster i's potential outcomes is $y_{i1} \in [-w_{i1}, w_{i1}]$ and $y_{i0} \in [-w_{i0}, w_{i0}]$; then the constraint in (1) requires $y_i \in [-w_i, w_i]$ for each cluster i, with $w_i = \sqrt{q(1-q)} \cdot \left(\frac{w_{i1}}{q} + \frac{w_{i0}}{1-q}\right)$. In this case, we claim that the optimal experiment is to simply assign treatment to each cluster independently with probability q. The correlation matrix with such independent assignment is the identity matrix, i.e., $\Sigma^* = I$. The worst-case potential outcome is $y_i = w_i$ for each cluster i, and the worst-case variance is $y^{\mathrm{T}}\Sigma^* y = \sum_{i\in[n]} w_i^2$.

We now show no experiment can achieve a worst-case variance strictly smaller than $\sum_{i\in[n]} w_i^2$. To see this, consider any feasible experiment and let Σ denote the corresponding correlation matrix. Consider the following randomized potential outcomes with Y_i being either w_i or $-w_i$ with equal probability, and let these Y_i be independent. Then, $\mathbb{E}\big[Y_i^2\big] = w_i^2$ and $\mathbb{E}\big[Y_i Y_k\big] = 0$ for any $i \neq k$. Thus, letting $Y = (Y_i)_{i\in[n]} \in \mathbb{R}^n$ be the concatenation, we have

$$\max_{y\in\times_{i\in[n]}[0,w_i]} y^{\mathrm{T}}\Sigma y \ge \mathbb{E}\Big[Y^{\mathrm{T}}\Sigma Y\Big] = \sum_{i\in[n]} w_i^2,$$

which clearly shows the optimality of independently assigning treatments to each cluster.

B.3 Proof of Proposition 3.1

Since the set of joint assignment distributions $\mathcal{P}_q$ is a polyhedron and $\Sigma(P)$ is a linear map of a distribution $P \in \mathcal{P}_q$, we can write (1) as a linear program with exponentially many decision variables and constraints as in (6), i.e.,

$$\begin{aligned} &\underset{P \in \mathcal{P}_q,\, z \in \mathbb{R}}{\text{minimize}} && z \\ &\text{subject to} && z \geq y^{\mathrm{T}} \Sigma(P) y,\ \forall\, y \in \bigtimes_{i \in [n]} \{0, w_i\}, \end{aligned} \tag{6}$$

where we have polyhedral constraints of $\mathcal{P}_q$ and we have one constraint for each extreme point of the potential outcomes' uncertainty set. This formulation also implies that a convex combination of optimal experiments is an optimal experiment as well.

Since clusters have equal sizes, this observation implies that given any optimal experiment, we can always construct another optimal experiment that treats clusters in an identical way. As a result, it is without loss of generality to assume that $\sigma_{ik} = \sigma$ for any two clusters. Thus, the variance of the estimator with a given set of potential outcomes becomes

$$y^{\mathrm{T}} \Sigma y = \sum_{i \in [n]} y_i^2 + \sigma \sum_{i \in [n]} \sum_{[n] \ni k \neq i} y_i y_k.$$

Since all of the potential outcomes are nonnegative, the optimal experiment tries to make σ as small as possible.

Let $S = \sum_{i \in [n]} Z_i$ denote the total number of clusters that receive treatments. We have

$$\mathbb{V}\text{ar}[S] = q(1-q) \Big[n + n(n-1)\sigma \Big],$$

simply because $\mathbb{C}\text{ov}\left[Z_i, Z_j\right] = q(1-q)\sigma$ for any two clusters $i \neq j$. Thus, to minimize the correlation σ, it is equivalent to minimizing the variance of the summation S. Since the mean value $\mathbb{E}[S] = qn$ is fixed, to minimize the variance $\mathbb{V}\text{ar}[S]$, the (exchangeable) joint assignment should let the summation S concentrate around the mean as much as possible.[14]

Case One: Suppose that qn is an integer. Since $\mathbb{V}\text{ar}[S] \geq 0$, we have $\sigma \geq -\frac{1}{n-1}$. The equality is attained when S is constant with probability one, and this can be achieved by assigning qn clusters to treatment uniformly at random. We next study the worst-case potential outcomes and the worst-case variance. Specifically, let $h \in \mathbb{N}$ be the number of clusters that have an outcome

[14]Similar problems are studied under more general marginal distributions (i.e., beyond Bernoulli distributions) in Rüschendorf and Uckelmann (2002) and Section 3.6 of Rachev and Rüschendorf (1998).

$y_i = 1$, and suppose that the other clusters have the outcome $y_i = 0$. Then,

$$V^{\text{OPT}} = \max_{h \in [0:n]} \; h + h(h-1)\sigma. \tag{7}$$

Denote by $h^* \in [0:n]$ the integer that maximizes this quantity; it is easy to see that h^* is the integer closest to $-\frac{1}{2\sigma} + \frac{1}{2} = \frac{n}{2}$. Hence, when n is even, we have $h^* = \frac{n}{2}$ and $V^{\text{OPT}} = \frac{1}{4}\frac{n^2}{n-1}$; and when n is odd, h^* is either $\frac{n+1}{2}$ or $\frac{n-1}{2}$, and $V^{\text{OPT}} = \frac{n+1}{4}$.

Case Two: Suppose that qn is not an integer. Let $p = \mathbb{P}[S < qn]$, $\underline{s} = \mathbb{E}[S|S < qn]$, and $\bar{s} = \mathbb{E}[S|S > qn]$. Since $\mathbb{E}[S] = qn$, by the law of total expectation, we have

$$p \cdot (qn - \underline{s}) = (1-p) \cdot (\bar{s} - qn). \tag{8}$$

Moreover, by the law of total variance,

$$\mathbb{V}\text{ar}[S] \geq p(qn - \underline{s})^2 + (1-p)(\bar{s} - qn)^2 = (qn - \underline{s})(\bar{s} - qn), \tag{9}$$

where the equality follows from (8), and an equality is attained at the inequality if S is a constant conditioning on $S < qn$ and $S > qn$, respectively.

Since the number of clusters in treatment S takes integral values, we have $\underline{s} \leq \lfloor qn \rfloor$ and $\bar{s} \geq \lceil qn \rceil$. As a result, from (9), the optimal experiment picks $\underline{s} = \lfloor qn \rfloor$ and $\bar{s} = \lceil qn \rceil = \underline{s}+1$, and it uniformly at random treats $\underline{s}$ clusters (and thus let $S = \underline{s}$) with probability p and uniformly at random treats $\bar{s}$ units (and thus let $S = \bar{s}$) with probability $1-p$. Moreover, by (8), $p = \frac{\bar{s}-qn}{\bar{s}-\underline{s}} = \lceil qn \rceil - qn$. The correlation of any two assignments is

$$\sigma = \frac{p \cdot \frac{\underline{s}}{n} \cdot \frac{\underline{s}-1}{n-1} + (1-p) \cdot \frac{\bar{s}}{n} \cdot \frac{\bar{s}-1}{n-1} - q^2}{q(1-q)} = -\frac{nq(1-q) - p(1-p)}{n(n-1)q(1-q)}. \tag{10}$$

To see that $\sigma < 0$, it suffices to show that $nq(1-q) - p(1-p) > 0$. First, if $qn < 1$, then $\underline{s} = 0$, $\bar{s} = 1$, and $p = 1 - qn$; thus, $nq(1-q) - p(1-p) = n(n-1)q^2 > 0$. On the other hand, suppose that $qn > 1$. Then, since $qn > 1 > \max\{p, 1-p\}$ and $1 - q \geq \frac{1}{2} \geq \min\{p, 1-p\}$, we again have $nq(1-q) > p(1-p)$. The fact that $\sigma > -\frac{1}{n-1}$ follows from (10) as $p = \lceil qn \rceil - qn \in (0,1)$. Also note that in (10), $\sigma = -\frac{1}{n-1}$ if $p = \lceil qn \rceil - qn = 0$, which happens when qn is an integer; thus, (10) also covers Case One.

Finally, let h^* denote the integer closest to $\min\left\{-\frac{1}{2\sigma} + \frac{1}{2}, n\right\}$. By (7), in the worst case, h^*

clusters have the outcome $y_i = 1$ and the other clusters have the outcome $y_i = 0$. When $n \to \infty$, we have $\sigma \to -\frac{1}{n-1}$ by (10) and therefore following a similar analysis to Case One, $V^{\mathrm{OPT}} \to \frac{n}{4}$.

B.4 Proof of Lemma 3.3

Since the objective of $\max_{y_i \in [0, w_i]} y^{\mathrm{T}} \Sigma y$ is jointly convex in y, in the worst case, $y_i \in \{0, w_i\}$ for each cluster $i \in [k]$. We first prove that there exists a worst-case potential outcome y such that $y_i = w_i$ for $i \le r$ for some integer $r \in [k]$, and $y_i = 0$ for $i > r$. If not, then for any worst-case potential outcome y, there exists indices $i < j$ such that $y_i = 0$ and $y_j = w_j$, whereas $w_i \ge w_j$. Since all the off-diagonals of the correlation matrix Σ have the same value of σ, the objective does not change if we instead swap y_i and y_j, and let $y_i = w_j \in [0, w_i]$ and $y_j = 0$. We can further (weakly) increase the objective by setting y_i to one of its extreme values, i.e., $y_i = w_i$ or $y_i = 0$. By iterating this process, we end up with a worst-case outcome that satisfies our desired property.

It remains to determine the value of r. Let y denote the worst-case potential outcome vector such that $y_i = w_i$ for $i \le r$ and $y_i = 0$ for $i > r$. Note that

$$y^{\mathrm{T}} \Sigma y = \sum_{i \in [k]} y_i^2 + \sigma \cdot \sum_{i \in [k]} \sum_{[k] \ni j \neq i} y_i y_j.$$

Observe that if we update $y_i = 0$ for some cluster $i \le r$, the variance changes by

$$-w_i^2 - 2\sigma \sum_{j \le r \text{ and } j \neq i} w_i w_j \le 0.$$

Similarly, if we set $y_i = w_i$ for some cluster $i > r$, the variance changes by

$$w_i^2 + 2\sigma \sum_{j=1}^{r} w_i w_j \le 0.$$

Together these imply that

$$\forall\, i \le r : \qquad w_i \ge -2\sigma \sum_{j \le r \text{ and } j \neq i} w_j, \tag{11}$$

$$\forall\, i > r : \qquad w_i \le -2\sigma \sum_{j=1}^{r} w_j. \tag{12}$$

Let $r^\star$ be the largest index such that $w_{r^\star} \ge -2\sigma \sum_{i \le r^\star - 1} w_i$. Since the cluster sizes are decreasing, $r^\star$ satisfies (11) and (12). Moreover, if $w_{r^\star} > -2\sigma \sum_{i \le r^\star - 1} w_i$, $r = r^\star$ is the only integer that satisfies both (11) and (12), and hence corresponds to a worst-case potential outcome. If $w_{r^\star} = -2\sigma \sum_{i \le r^\star - 1} w_i$, then both $r = r^\star$ and $r = r^\star - 1$ satisfy (11) and (12), and they both constitute a worst-case potential outcome.

B.5 Proof of Lemma 4.2

The lower bound $w_1^2 \le V^{\text{LB}}$ in Lemma 4.2 is trivial because $y^\top \Sigma y = w_1^2$ with the outcome vector y such that $y_1 = w_1$ and $y_i = 0$ for all $i \ge 2$, for any correlation matrix Σ. Thus, we only need to prove the inequality $\frac{1}{4} \sum_{i \in [n]} w_i^2 \le V^{\text{LB}}$.

To prove this, consider the following randomized potential outcomes where each Y_i is either 0 or w_i with equal probability, and these Y_i are independent. Then, we have $\mathbb{E}\big[Y_i^2\big] = w_i^2/2$ and $\mathbb{E}\big[Y_i Y_k\big] = w_i w_k/4$ for any $i \ne k$. Let $Y = (Y_i)_{i \in [n]}$ be the concatenation of these randomized potential outcomes and $w = (w_i)_{i \in [n]}$ be the vector of cluster sizes. For any correlation matrix $\Sigma = (\sigma_{ik})_{i,k \in [n]} \in \Re$, we have

$$\max_{y \in \times_{i \in [n]} [0, w_i]} y^\top \Sigma y \ge \mathbb{E}\Big[Y^\top \Sigma Y\Big] = \frac{1}{2} \sum_{i \in [n]} w_i^2 + \frac{1}{4} \sum_{i \in [n]} \sum_{[n] \ni k \ne i} w_i w_k \sigma_{ik} = \frac{1}{4} \sum_{i \in [n]} w_i^2 + \frac{1}{4} w^\top \Sigma w \ge \frac{1}{4} \sum_{i \in [n]} w_i^2.$$

Thus, $V^{\text{LB}} \ge \frac{1}{4} \sum_{i \in [n]} w_i^2$, which completes the proof of the lemma.

B.6 Proof of Lemma 4.3

We start by finding an upper bound on the worst-case variance within a group. First, it is clear that the worst-case variance increases with the size w_i of any cluster i in the group because increasing w_i enlarges the potential outcomes' uncertainty set. Thus, for a group with at most k clusters where the largest cluster size is equal to w, the worst-case variance is largest when there are exactly k clusters and all cluster sizes are equal to w. By Proposition 3.1, the optimal partition is to have all the k clusters in one block, and the corresponding worst-case variance is $f_q(k) \cdot w^2$.

Next, let $N = \lceil \frac{n}{k} \rceil$ be the number of groups. By the above analysis, we have

$$V^k \le f_q(k) S_k, \tag{13}$$

where $S_k = \sum_{i=1}^N w_{(i-1)k+1}^2$ is the sum of squares of the largest cluster sizes in each group. We claim that

$$4V^{\text{LB}} \ge \sum_{i \in [n]} w_i^2 \ge k \cdot (S_k - w_1^2) + w_1^2. \tag{14}$$

The first inequality in (14) follows from Lemma 4.2. The second inequality in (14) holds, because

$$\sum_{i\in[n]} w_i^2 \geq w_1^2 + \sum_{i=2}^{(N-1)k+1} w_i^2 = w_1^2 + \sum_{h=1}^{N-1} \sum_{i=(h-1)k+2}^{hk+1} w_i^2$$
$$\overset{(*)}{\geq} w_1^2 + k\cdot\sum_{h=1}^{N-1} w_{hk+1}^2 = w_1^2 + k\cdot(S_k - w_1^2),$$

where the above inequality $(*)$ follows from the fact that clusters $(h-1)k+2$ to hk have weakly larger sizes than the cluster $hk+1$. Now, by rearranging the terms in (14), we have

$$S_k \leq \frac{4V^{\text{LB}}}{k} + \frac{k-1}{k} w_1^2.$$

Combining this with (13) yields

$$\frac{V^k}{V^{\text{LB}}} \leq f_q(k)\cdot\frac{S_k}{V^{\text{LB}}} \leq f_q(k)\left(\frac{4}{k} + \frac{k-1}{k}\frac{w_1^2}{V^{\text{LB}}}\right) \leq f_q(k)\left(\frac{4}{k} + \frac{k-1}{k}\right),$$

where the last inequality follows from $w_1^2 \leq V^{\text{LB}}$ by Lemma 4.2.

B.7 Proof of Lemma 4.7

It suffices to prove Lemma 4.7 only for $k \geq k_0 \triangleq \frac{1}{2q(1-q)} \geq 2$, because when $k \leq k_0$, $f_q(k)$ can be uniformly bounded from above by a constant. By Proposition 3.1, $\sigma = -\frac{nq(1-q)-p(1-p)}{n(n-1)q(1-q)} < 0$ with $p = \lceil qn \rceil - qn$; hence, we have

$$f_q(k) = \max_{h\in[k]}\{h + h(h-1)\sigma\} \leq \max_{h\in\mathbb{R}}\{h + h(h-1)\sigma\} = \frac{k^3 q(1-q)}{4(k-1)\big[kq(1-q) - p(1-p)\big]},$$

where the second equality follows by taking $h = -\frac{1}{2\sigma} + \frac{1}{2}$ that maximizes the quadratic objective. Since $p(1-p) \leq \frac{1}{4} = \frac{k_0}{2}\cdot q(1-q)$, we have

$$4f_q(k) - k \leq \frac{k^3}{(k-1)\big(k-\frac{k_0}{2}\big)} - k = \frac{k^2\big(1+\frac{k_0}{2}\big) - \frac{k_0}{2}k}{(k-1)\big(k-\frac{k_0}{2}\big)} \leq \left(1+\frac{k_0}{2}\right)\frac{1}{1-\frac{1}{k}}\frac{1}{1-\frac{k_0}{2k}}.$$

By the fact that $\frac{1}{1-x} \leq 1 + 2x$ for $x \in [0, \frac{1}{2}]$ and that $k \geq k_0 \geq 2$, we have

$$4f_q(k) - k \leq \left(1+\frac{k_0}{2}\right)\left(1+\frac{2}{k}\right)\left(1+\frac{k_0}{k}\right) \leq 2\left(1+\frac{k_0}{2}\right)\left(1+\frac{2}{k_0}\right) = 4 + 8q(1-q) + \frac{1}{2q(1-q)},$$

which is bounded from above by a constant.

B.8 Details and Proof of Correctness of Example 4.1

Throughout this section, we focus on the setting of Example 4.1. For notational convenience, we index the blocks of an IBR experiment in decreasing order of size from the largest cluster in each block. Also, for simplicity, we assume an even number n of clusters. We first show in Lemma B.1 that when a block contains four or more clusters, at least $p \geq 2$ clusters take positive values (i.e., non-zero outcomes) in the worst-case potential outcome.

Lemma B.1. *Suppose that a block contains an even number $k \geq 4$ clusters, with cluster sizes $w_i = \beta^{k-i} w_k$ for any $i \in [k]$. Then, at least $p \geq 2$ clusters take positive values in the worst-case potential outcome $y = (y_i)_{i\in[k]}$.*

Proof. Since the number of clusters k is even, the correlation between any two clusters is $\sigma = -\frac{1}{k-1}$ by Corollary 3.2. By Lemma 3.3, p is the largest integer that satisfies

$$\beta^{k-p} w_k = w_p \geq \frac{2}{k-1} \cdot \sum_{i=1}^{p-1} w_i = \frac{2}{k-1} \cdot \sum_{i=1}^{p-1} \beta^{k-i} w_k = \frac{2}{k-1} \frac{\beta^p - \beta}{\beta - 1} \cdot \beta^{k-p} w_k.$$

This implies that $p = \left\lfloor \frac{\ln\left(\frac{k-1}{2}(\beta-1)+\beta\right)}{\ln \beta} \right\rfloor \geq 2$ when $k \geq 4$. $\square$

We next show in Lemma B.2 that with the optimal partition, all blocks contain either two or four clusters.

Lemma B.2. *Suppose that the number of clusters n is even. All blocks in an optimal partition contain either 2 or 4 clusters.*

Proof. By Lemma C.3, each block contains an even number of clusters. Suppose that a block instead contains k clusters where k is an even number satisfying $k \geq 6$. Without loss of generality, we assume that $w_i = \beta^{k-i}$ for each cluster $i \in [k]$ (since we can always normalize cluster sizes by the size of the smallest cluster). We claim that we can further partition this block into two smaller blocks to reduce the worst-case variance. Specifically, the first block contains the first $k-2$ clusters of the original block, and the second block contains the other two blocks.

First, consider the worst-case variance associated with the two new blocks, denoted by V_a. Let p be the number of positive values in the worst-case potential outcome of the first block. By Lemma B.1, $p \geq 2$ because block one contains $k-2 \geq 4$ clusters. We have

$$V_a = \left(\sum_{i\in[p]} w_i^2 - \sum_{i\in[p]} \sum_{[p]\ni j\neq i} \frac{1}{k-3} w_i w_j \right) + \beta^2.$$

Here β^2 is simply the worst-case variance of the second block, because in the worst case, cluster $k-1$ has a positive outcome and cluster k has outcome zero (this follows from Lemma C.2).

Let V_b denote the worst-case variance of the original block. It satisfies

$$V_b \geq \sum_{i\in[p]} w_i^2 - \sum_{i\in[p]} \sum_{[p]\ni j\neq i} \frac{1}{k-1} w_i w_j,$$

because the correlation between any two clusters in the original block is $-\frac{1}{k-1}$, and p is not necessarily the number of positive values in the worst-case potential outcome of the original block. Thus,

$$\begin{aligned} V_b - V_a &\geq \sum_{i\in[p]} \sum_{[p]\ni j\neq i} \left(\frac{1}{k-3} - \frac{1}{k-1}\right) w_i w_j - \beta^2 \\ &\geq 2\left(\frac{1}{k-3} - \frac{1}{k-1}\right) w_1 w_2 - \beta^2 \\ &= \frac{4\beta^{2k-3}}{(k-1)(k-3)} - \beta^2, \end{aligned}$$

which is nonnegative when $k \geq 6$. Hence, splitting the large block into two smaller blocks reduces the worst-case variance. □

Finally, we show in Lemma B.3 that all blocks contain four clusters in the optimal partition.

Lemma B.3. *Suppose that the number of clusters n is even. Then the optimal partition of an IBR experiment satisfies the following:*

- *If n is divisible by 4, all blocks contain exactly 4 clusters;*
- *Otherwise, the last block contains 2 clusters, and all the other blocks contain exactly 4 clusters.*

Proof. By Lemma B.2, each block contains either two or four clusters. For a two-cluster block, the worst-case outcome is simply the large cluster having a positive outcome and the small cluster having the zero outcome. For a four-cluster block, by the proof of Lemma B.1, the worst-case potential outcome is only the two largest clusters taking positive outcomes.

Let K be the number of blocks. It suffices to show that all of the first $K-1$ blocks contain four clusters. Suppose by way of a contradiction that block $h \leq K-1$ contains two clusters, k and $k+1$. If block $h+1$ contains two clusters $(k+2,\, k+3)$ as well, then the above observations imply that merging blocks h and $h+1$ decreases the worst-case variance by

$$w_{k+3}^2 \cdot \left\{ \left(\beta^6 + \beta^2\right) - \left(\beta^6 + \beta^4 - \frac{2}{3} \cdot \beta^5\right) \right\} > 0.$$

Now suppose that block $h+1$ contains four clusters. Suppose that we reconstruct blocks h and $h+1$ by assigning clusters k to $k+3$ to block h and clusters $k+4$ and $k+5$ to block $h+1$. Then, the worst-case variance decreases by

$$w_{k+5}^2 \cdot \left\{ \left(\beta^{10} + \beta^6 + \beta^4 - \frac{2}{3} \cdot \beta^5\right) - \left(\beta^{10} + \beta^8 - \frac{2}{3} \cdot \beta^9 + \beta^2\right) \right\} > 0.$$

Thus, it follows that it is not optimal for any block $h \leq K-1$ to contain two clusters. $\square$

We now show that the optimal IBR experiment is asymptotically suboptimal. Consider a block with four clusters with sizes $w_i = \beta^{4-i}$ for $i \in [4]$. The worst-case variance from randomly assigning half of the clusters to treatment is $v^{\text{half}} = \beta^6 + \beta^4 - \frac{2}{3} \cdot \beta^5 = 4.222$. The worst-case variance from the optimal randomized joint assignment (given in Table 3) is $v^{\text{OPT}} = 3.815$. Now, let us revisit Example 4.1. For simplicity, we assume that the number of clusters n is divisible by four. Since every block in the optimal partition contains 4 clusters by Lemma B.3 and the blocks are identical up to a scaling, the worst-case variance from the optimal IBR experiment is

$$V^{\text{DP}} = \sum_{i=1}^{n/4} w_{4i-3}^2 \cdot v^{\text{half}}.$$

For an experiment that assigns clusters in a block to treatment in an optimal way, and does so independently across blocks, the worst-case variance, denoted by V, is

$$V = \sum_{i=1}^{n/4} w_{4i-3}^2 \cdot v^{\text{OPT}}.$$

Thus,

$$\frac{V^{\text{DP}} - V^{\text{OPT}}}{V^{\text{OPT}}} \geq \frac{V^{\text{DP}} - V}{V} = \frac{v^{\text{half}} - v^{\text{OPT}}}{v^{\text{OPT}}} = 10.7\%,$$

which implies that the optimal IBR experiment is asymptotically strictly suboptimal.

	1	2	3	4	probability
1	$\surd$	$\times$	$\surd$	$\times$	0.1700
2	$\times$	$\surd$	$\times$	$\surd$	0.1700
3	$\surd$	$\surd$	$\times$	$\times$	0.1500
4	$\times$	$\times$	$\surd$	$\surd$	0.1500
5	$\surd$	$\times$	$\times$	$\surd$	0.0954
6	$\times$	$\surd$	$\surd$	$\times$	0.0954
7	$\surd$	$\times$	$\times$	$\times$	0.0846
8	$\times$	$\surd$	$\surd$	$\surd$	0.0846

Table 3: The optimal randomized joint assignment of treatment to four clusters, with cluster sizes $w_i = \beta^{4-i}$ for $i \leq 4$ and $\beta = \frac{5}{4}$. Each of the 8 rows corresponds to an assignment, where $\surd$ denotes treatment and $\times$ denotes control.

B.9 Proof of Theorem 4.9

Throughout the proof, we refer to a cluster of size w_k as a cluster of type k. First, using a similar argument to the one in the proof of Proposition 3.1, we have that any convex combination of optimal correlation matrices in (5) (which is the optimization problem that defines V^{LB}) is an optimal correlation matrix as well. Thus, there exists an optimal correlation matrix attaining V^{LB} that takes the following form:

$$\Sigma = \begin{pmatrix} \sigma_{11}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{11})I & \sigma_{12}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \cdots & \sigma_{1K}\mathbf{1}\mathbf{1}^{\mathrm{T}} \\ \sigma_{12}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \sigma_{22}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{22})I & & \sigma_{2K}\mathbf{1}\mathbf{1}^{\mathrm{T}} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \sigma_{2K}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \cdots & \sigma_{KK}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{KK})I \end{pmatrix}, \tag{15}$$

where σ_{kk} is the correlation coefficient of the treatment assignments for any two different clusters of the same type k, and $\sigma_{k\ell}$ is the correlation coefficient of the treatment assignments for any two clusters of types k and ℓ with $k \neq \ell$, respectively. Since the correlation matrix Σ needs to be positive semidefinite, for any scalars $x_k \in \mathbb{R}$ for $k \in [K]$, we have

$$
\begin{aligned}
0 &\le \begin{pmatrix} x_1 \mathbf{1}^{\mathrm{T}} & x_2 \mathbf{1}^{\mathrm{T}} & \cdots & x_K \mathbf{1}^{\mathrm{T}} \end{pmatrix} \Sigma \begin{pmatrix} x_1 \mathbf{1} \\ x_2 \mathbf{1} \\ \vdots \\ x_K \mathbf{1} \end{pmatrix} \\
&= \sum_{k=1}^{K} \left[\sigma_{kk} x_k^2 n_k^2 + (1 - \sigma_{kk}) x_k^2 n_k \right] + \sum_{k=1}^{K} \sum_{\ell \neq k} \sigma_{k\ell} x_k x_\ell n_k n_\ell \\
&= \begin{pmatrix} x_1 n_1 & x_2 n_2 & \cdots & x_K n_K \end{pmatrix} \begin{pmatrix} \frac{1+(n_1-1)\sigma_{11}}{n_1} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{12} & \frac{1+(n_2-1)\sigma_{22}}{n_2} & & \sigma_{2K} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \cdots & \frac{1+(n_K-1)\sigma_{KK}}{n_K} \end{pmatrix} \begin{pmatrix} x_1 n_1 \\ x_2 n_2 \\ \vdots \\ x_K n_K \end{pmatrix}.
\end{aligned}
$$

Thus, the matrix

$$
\tilde{\Sigma} = \begin{pmatrix} \frac{1+(n_1-1)\sigma_{11}}{n_1} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{12} & \frac{1+(n_2-1)\sigma_{22}}{n_2} & & \sigma_{2K} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \cdots & \frac{1+(n_K-1)\sigma_{KK}}{n_K} \end{pmatrix} \succeq 0 \tag{16}
$$

needs to be positive semidefinite.

For the IBR experiment with one block for each set of clusters of the same type, the corresponding correlation matrix takes the form of (15) as well, with σ_{kk} as given in Proposition 3.1 and $\sigma_{k\ell} = 0$ for all $k \neq \ell$. For each block k, let $\bar{h}_k$ denote the number of clusters in the block that take values w_k in the worst-case potential outcome of the IBR experiment (the other k-type clusters take value 0). Now, let Σ^* denote an optimal correlation matrix attaining V^{LB} that takes the form of (15) and let $(\sigma^*_{k\ell})_{k,\ell \in [K]}$ denote the corresponding correlation coefficients between cluster types. We have

$$
\begin{aligned}
V^{\mathrm{LB}} &= \max_{h_k \in [0:n_k]} \left\{ \sum_{k \in [K]} w_k^2 \left[h_k + h_k (h_k - 1) \sigma^*_{kk} \right] + \sum_{k \in [K]} \sum_{\ell \neq k} w_k w_\ell h_k h_\ell \sigma^*_{k\ell} \right\} \\
&\ge \sum_{k \in [K]} w_k^2 \left[\bar{h}_k + \bar{h}_k (\bar{h}_k - 1) \sigma^*_{kk} \right] + \sum_{k \in [K]} \sum_{\ell \neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma^*_{k\ell},
\end{aligned}
$$

where we plug in the worst-case potential outcome of the IBR experiment to obtain the inequality.

As a result,

$$
\begin{aligned}
V - V^{\mathrm{LB}} &\le \sum_{k\in[K]} w_k^2 \bar{h}_k \left(\bar{h}_k - 1\right)\left(\sigma_{kk} - \sigma_{kk}^*\right) - \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma_{k\ell}^* \\
&= \sum_{k\in[K]} w_k^2 \left(\bar{h}_k\left(\bar{h}_k - 1\right)\left(\sigma_{kk} - \sigma_{kk}^*\right) + \frac{1+(n_k-1)\sigma_{kk}^*}{n_k}\bar{h}_k^2\right) \\
&\quad - \left\{\sum_{k\in[K]} \frac{1+(n_k-1)\sigma_{kk}^*}{n_k} w_k^2 \bar{h}_k^2 + \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma_{k\ell}^*\right\} \\
&\le \sum_{k\in[K]} w_k^2 \left(\bar{h}_k\left(\bar{h}_k - 1\right)\left(\sigma_{kk} - \sigma_{kk}^*\right) + \frac{1+(n_k-1)\sigma_{kk}^*}{n_k}\bar{h}_k^2\right) \qquad (17) \\
&= \sum_{k\in[K]} w_k^2 \left[\left(\bar{h}_k - \frac{\bar{h}_k^2}{n_k}\right)\sigma_{kk}^* + \left(\underbrace{\bar{h}_k^2\left(\frac{1}{n_k} + \sigma_{kk}\right) - \bar{h}_k\sigma_{kk}}_{(a)}\right)\right] \\
&\le \sum_{k\in[K]} w_k^2 \left(\frac{1}{4} n_k \sigma_{kk}^* + \frac{1}{2q(1-q)} + 1\right),
\end{aligned}
$$

where the second inequality follows from the fact that with $u = \left(w_k \bar{h}_k\right)_{k\in[K]} \in \mathbb{R}^K$,

$$
\sum_{k\in[K]} \frac{1+(n_k-1)\sigma_{kk}^*}{n_k} w_k^2 \bar{h}_k^2 + \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma_{k\ell}^* = u^{\mathrm{T}}\tilde{\Sigma} u \ge 0,
$$

because $\tilde{\Sigma}$ is positive semidefinite by (16), and the last inequality follows from the fact that $\bar{h}_k - \frac{\bar{h}_k^2}{n_k} \le \frac{n_k}{4}$ (the equality is attained when $\bar{h}_k = \frac{n_k}{2}$) and that $(a) \le \frac{1}{2q(1-q)} + 1$ by Lemma B.4. To bound the value of $n_k \sigma_{kk}^*$ from above, note that by Lemma 4.2,

$$
V^{\mathrm{LB}} \le \sum_{k\in[K]} n_k w_k^2 \le n w_1^2. \qquad (18)
$$

Also, note that

$$
\begin{aligned}
V^{\mathrm{LB}} &= \max_{h_k\in[0:n_k]} \left\{\sum_{k\in[K]} w_k^2 \Big[h_k + h_k(h_k - 1)\sigma_{kk}^*\Big] + \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell h_k h_\ell \sigma_{k\ell}^*\right\} \\
&\ge w_k^2 \Big[n_k + n_k(n_k-1)\sigma_{kk}^*\Big] \qquad (19) \\
&\ge w_k^2 n_k^2 \sigma_{kk}^*,
\end{aligned}
$$

where we take $h_k = n_k$ and $h_\ell = 0$ for all $\ell \neq k$ to obtain the first inequality. Combining the inequalities (18) and (19) gives

$$n_k \sigma^*_{kk} \le \min\left\{ n_k, \frac{w_1^2 n}{w_k^2 n_k} \right\}.$$

Plugging this last inequality into the right-hand side of the chain of inequalities in (17) yields the desired bound, and hence completes the proof.

Lemma B.4. $(a) \le \frac{1}{2q(1-q)} + 1$.

Proof. First, if $n_k = 1$, we have $\bar{h}_k = 1$ and hence $(a) = 1 \le \frac{1}{2q(1-q)} + 1$. In what follows, we assume that $n_k \ge 2$.

Note that $\sigma_{kk} = -\frac{n_k q(1-q) - p(1-p)}{n_k(n_k-1)q(1-q)} \in [-\frac{1}{n_k-1}, 0)$ with $p = \lceil q n_k \rceil - q n_k$ by Proposition 3.1. If $p(1-p) \le q(1-q)$, then we have $\sigma \le -\frac{1}{n_k}$ and, as a result, $(a) \le -\bar{h}_k \sigma_{kk} \le \frac{n_k}{n_k - 1} \le 2 \le \frac{1}{2q(1-q)} + 1$.

Otherwise, if $p(1-p) \ge q(1-q)$, then we have $\sigma_{kk} \ge -\frac{1}{n_k}$ and $\sigma_{kk} + \frac{1}{n_k} = \frac{p(1-p) - q(1-q)}{n_k(n_k-1)q(1-q)} \le \frac{1}{4n_k(n_k-1)q(1-q)}$. As a result,

$$(a) \le n_k^2 \cdot \frac{1}{4n_k(n_k-1)q(1-q)} + n_k \cdot \frac{1}{n_k} = \frac{n_k}{n_k - 1} \frac{1}{4q(1-q)} + 1 \le \frac{1}{2q(1-q)} + 1. \qquad \square$$

B.10 Proof of Theorem 4.10

First, note that we have

$$K \le \left\lceil \frac{\ln w_1/\bar{w}}{\ln \alpha} \right\rceil \le \frac{\ln w_1/\bar{w}}{\ln \alpha} + 1 \le \frac{(1+\delta_1)\ln n}{2\ln(1+n^{-\delta_2})} + 1 = O\left(n^{\delta_2} \cdot \ln n\right).$$

Let $S_k \subseteq [n]$ denote the set of clusters in block k and let $n_k = |S_k|$ denote this block's cardinality. Consider a new problem instance in which we first drop block zero, and then for each block $k \in [K]$ we decrease all of the cluster sizes to the smallest cluster size in the block. We denote the smallest cluster size in the block k by $\underline{w}_k \triangleq \min_{i \in S_k} w_i$. For this new problem instance, let $\tilde{V}^{\text{LB}}$ denote the lower bound on the worst-case variance of an optimal experiment (i.e., the optimal value of (5)) and let $\tilde{V}$ denote the worst-case variance of the IBR experiment. We proceed by bounding the differences of the original and the new problem instances in terms of the worst-case variances of their IBR experiments and the corresponding lower bounds.

Lemma B.5. $\tilde{V}^{\text{LB}} \le V^{\text{LB}}$ *and* $0 \le V - \tilde{V} \le n^{-\delta_1} \sum_{i=1}^n w_i^2 + (\alpha^2 - 1) \sum_{i=1}^n w_i^2$.

Proof. Since cluster sizes are weakly smaller in the new problem instance, the set of the potential outcomes is more restricted, and hence $\tilde{V}^{\text{LB}} \le V^{\text{LB}}$ and $\tilde{V} \le V$. For an IBR experiment, the worst-case variance is the sum of the worst-case variances of each block. Let V_k and $\tilde{V}_k$ denote the worst-case variances of block k in the original and the new problem instances, respectively. We have

$$
\begin{aligned}
V - \tilde{V} &= V_0 + \sum_{k\in[K]} \left(V_k - \tilde{V}_k\right) \\
&\le \sum_{i\in S_0} w_i^2 + \sum_{k\in[K]} \left(\alpha^2 \tilde{V}_k - \tilde{V}_k\right) \\
&\le n^{-\delta_1} \sum_{i=1}^{n} w_i^2 + (\alpha^2-1)\tilde{V} \\
&\le n^{-\delta_1} \sum_{i=1}^{n} w_i^2 + (\alpha^2-1) \sum_{i=1}^{n} w_i^2.
\end{aligned}
\tag{20}
$$

Here, the first inequality uses the fact that for the specific logarithmic partition, we have $w_i \le \alpha \underline{w}_k$ for any cluster $i \in S_k$. This implies that the worst-case variance V_k is no larger than the worst-case variance when all cluster sizes in block k are equal to $\alpha \underline{w}_k$. The latter quantity is $\alpha^2 \tilde{V}_k$, and hence $V_k \le \alpha^2 \tilde{V}_k$. The first and third inequalities also make use of the inequalities $V_0 \le \sum_{i\in S_0} w_i^2$ and $\tilde{V} \le V \le \sum_{i=1}^n w_i^2$, respectively, by Lemma 4.2. These hold because an IBR experiment always has a smaller worst-case variance than the naive experiment that assigns treatment to each cluster independently. Finally, the second inequality follows from the fact that $w_i \le \bar{w}$ for all clusters $i \in S_0$ and that $n_0 = |S_0| \le n$. □

We next analyze the new problem instance, and bound the gap $\tilde{V} - \tilde{V}^{\text{LB}}$. In the new problem instance, clusters in a block are of equal size. Therefore, we can adopt the analysis for Theorem 4.9. Specifically, by (17) we have

$$
\tilde{V} - \tilde{V}^{\text{LB}} \le \frac{1}{4} \sum_{k\in[K]} \underline{w}_k^2 \cdot \left(n_k \sigma_{kk}^* + C(q)\right),
$$

where $C(q) \triangleq \frac{2}{q(1-q)} + 4$ is a constant and σ_{kk}^* is the optimal correlation coefficient of any two clusters in block k obtained from the solution of (5) (that takes the form (15)) for the new problem instance. By (19),

$$
\underline{w}_k^2 n_k \sigma_{kk}^* \le \min\left\{\underline{w}_k^2 n_k, \frac{\tilde{V}^{\text{LB}}}{n_k}\right\} \le \underline{w}_k \sqrt{\tilde{V}^{\text{LB}}} \le \underline{w}_k \sqrt{V^{\text{LB}}}.
$$

Thus, we have

$$\tilde{V} - \tilde{V}^{\text{LB}} \leq \frac{1}{4} \sum_{k \in [K]} \left\{ \underline{w}_k \sqrt{V^{\text{LB}}} + C(q) \cdot \underline{w}_k^2 \right\} \leq \frac{K}{4} \cdot \left(w_1 \sqrt{V^{\text{LB}}} + C(q) \cdot w_1^2 \right) \leq \frac{C(q)+1}{4} K \cdot w_1 \sqrt{V^{\text{LB}}}, \tag{21}$$

where the last inequality follows from $w_1^2 \leq V^{\text{LB}}$ by Lemma 4.2.

Combining (20) and (21), we have

$$\begin{aligned}
\frac{V - V^{\text{LB}}}{V^{\text{LB}}} &\leq \frac{V - \tilde{V} + \tilde{V} - \tilde{V}^{\text{LB}}}{V^{\text{LB}}} \\
&\leq \frac{4n^{-\delta_1} \sum_{i=1}^n w_i^2 + 4(\alpha^2 - 1) \sum_{i=1}^n w_i^2 + (C(q)+1)\, K \cdot w_1 \sqrt{V^{\text{LB}}}}{4V^{\text{LB}}} \\
&\leq \frac{4n^{-\delta_1} \sum_{i=1}^n w_i^2 + 4(\alpha^2 - 1) \sum_{i=1}^n w_i^2 + (C(q)+1)\, K \cdot w_1 \sqrt{\sum_{i=1}^n w_i^2}}{\sum_{i=1}^n w_i^2} \\
&= 4n^{-\delta_1} + 4(\alpha^2 - 1) + (C(q)+1)\, K \cdot \sqrt{\frac{w_1^2}{\sum_{i=1}^n w_i^2}} \\
&= O\left(n^{-\delta_1}\right) + O\left(n^{-\delta_2}\right) + O\left(n^{-\frac{c}{2}+\delta_2} \cdot \ln n\right),
\end{aligned}$$

where the third inequality follows from the lower and upper bounds of V^{LB} in Lemma 4.2. Taking $\delta_1 = \delta_2 = \frac{c}{4}$ results in

$$\frac{V - V^{\text{LB}}}{V^{\text{LB}}} = O\left(n^{-\frac{c}{4}} \ln n\right),$$

which completes the proof of the theorem.

C Structural Properties of IBR Experiments

We first show in Lemma C.1 that the correlation of any two clusters in a block is nondecreasing as the size of the block increases.

Lemma C.1. *Suppose that there are n clusters, and consider the experiment that selects a fraction q of the clusters uniformly at random and assigns them to treatment, as described in Proposition 3.1. The correlation coefficient σ of any two assignments is nondecreasing in the number of clusters n.*

Proof. Let σ_n denote the correlation coefficient of any two clusters when there are n clusters; by Proposition 3.1, we have $\sigma_n = -\frac{nq(1-q)-p_n(1-p_n)}{n(n-1)q(1-q)}$, with $p_n \triangleq \lceil qn \rceil - qn$.

To show that $\sigma_{n+1} \geq \sigma_n$, it is equivalent to show

$$(*) \triangleq (n+1)q(1-q) + (n-1)p_{n+1}(1-p_{n+1}) - (n+1)p_n(1-p_n) \geq 0.$$

In what follows, we check the four possible cases regarding the values of qn and $q(n+1)$ relative to $\lfloor qn \rfloor$ and $\lceil qn \rceil$ (as illustrated in Figure 4), and we validate the nonnegativity of $(*)$ for each case.

Case One: $qn \in \mathbb{N}$**.** In this case, $qn = \lfloor qn \rfloor = \lceil qn \rceil$ and hence, $p_n = 0$. Thus, $(*) > 0$ and hence $\sigma_{n+1} > \sigma_n$.

Case Two: $\lfloor qn \rfloor < qn < q(n+1) < \lceil qn \rceil$**.** In this case, $(*) = 2\lfloor qn \rfloor\big(\lceil qn \rceil - q(n+1)\big) \geq 0$ and hence $\sigma_{n+1} \geq \sigma_n$. Specifically, if furthermore $qn < 1$, $(*) = 0$ and hence $\sigma_{n+1} = \sigma_n$; otherwise, $\sigma_{n+1} > \sigma_n$.

Case Three: $\lfloor qn \rfloor < qn < q(n+1) = \lceil qn \rceil$**.** In this case, $p_{n+1} = 0$ and $p_n = q$. Hence, $(*) = 0$ and $\sigma_{n+1} = \sigma_n$.

Case Four: $\lfloor qn \rfloor < qn < \lceil qn \rceil < q(n+1)$**.** In this case, $(*) = 2\big(n - \lceil qn \rceil\big)\big((n+1)q - \lceil qn \rceil\big) > 0$; hence, $\sigma_{n+1} > \sigma_n$.

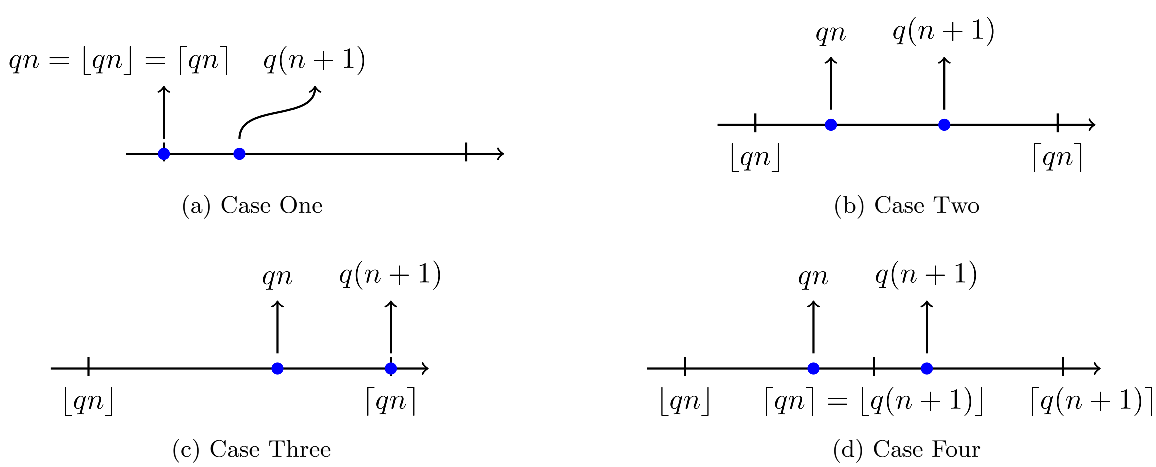


Figure 4: Illustration of the four cases.

□

In the rest of the section, we assume that the marginal assignment probability q is equal to $\frac{1}{2}$, and we shed light on two structural properties of the IBR experiments. Our first property establishes that under an IBR experiment, at most half of the clusters in a block can take a positive value in the worst-case potential outcome.

Lemma C.2. *Suppose that the marginal assignment probability q is equal to $\frac{1}{2}$. Consider a block with k clusters, and let r be the number of clusters that take a positive value in the worst-case potential outcome. Then, $r \leq \frac{k}{2}$ if k is even, and $r \leq \frac{k+1}{2}$ if k is odd.*

Proof. To see this, note that when the correlation σ is negative, Lemma 3.3 implies that

$$w_r \geq -2\sigma \sum_{i \leq r-1} w_i \geq -2(r-1)\sigma \cdot w_r,$$

which in turn implies that $r \leq -\frac{1}{2\sigma} + 1$. When the block contains an even number k of clusters, $\sigma = -\frac{1}{k-1}$ by Corollary 3.2. Thus, r is an integer no larger than $\frac{k}{2}$. Analogously, when the number of clusters k is odd, $\sigma = -\frac{1}{k}$, and hence, r is an integer no larger than $\frac{k+1}{2}$. $\square$

For the next result, we index the blocks of an IBR experiment in decreasing order of size from the largest cluster in the block (and we break ties using the size of the smallest cluster in the block). We show in Lemma C.3 that in the optimal partition obtained from solving the DP, all but the last block contain an even number of clusters.

Lemma C.3. *Suppose that the marginal assignment probability q is equal to $\frac{1}{2}$. There exists an optimal cluster partition obtained from solving the DP such that (i) if the number of clusters n is even, then all blocks contain an even number of clusters, and (ii) if n is odd, then all but the last block contain an even number of clusters.*

Proof. We first focus on a single block with k clusters and cluster sizes $w_1 \geq w_2 \geq \cdots \geq w_k$. We highlight three observations. First, clearly, if the vector of cluster sizes entry-wise decreases, the worst-case variance of this block will weakly decrease. Second, if we drop any of the clusters from the block, the worst-case variance of the block will weakly decrease as well because assignments to the remaining clusters become more negatively correlated. Third, if the number of clusters k is odd, adding a cluster of size $w' \leq w_k$ does not change the worst-case variance of this block. To see this, note that after the addition, the correlation between the assignments to any two clusters in the block does not change (and remains $-\frac{1}{k}$ by Corollary 3.2). Moreover, the worst-case potential outcome does not change either, due to Lemma 3.3 and Lemma C.2.

We now turn to the optimal partition by solving the DP. Let K denote the number of blocks, and $S_k = \left\{w_1^k \geq w_2^k \geq \cdots \geq w_{n_k}^k\right\}$ denote the set of clusters (sorted in decreasing order of size) in block $k \in [K]$. If a block $h \leq K-1$ has an odd number of clusters, consider a new partition $\{S'_k\}_{k\in[K]}$ as follows: $S'_k = S_k$ for any $k \leq h-1$, $S'_h = S_h \cup \left\{w_1^{h+1}\right\}$, $S'_k = S_k \setminus \{w_1^k\} \cup \left\{w_1^{k+1}\right\}$ for any $h+1 \leq k \leq K-1$, and $S'_K = S_K \setminus \{w_1^K\}$. By the former discussion, the worst-case variance of each block weakly decreases. Thus, $\{S'_k\}_{k\in[K]}$ is a weakly better partition than $\{S_k\}_{k\in[K]}$. Iterating this last step completes the proof. $\square$

D More on the Airbnb Example in Section 5.2

In this section, we elaborate more on the Airbnb example in Section 5.2.

D.1 Visualization and More Information about the Partition

We visualize the geographic locations of the listings in Figure 5. Figure 6 exhibits the ten largest clusters from the Louvain algorithm, and Table 4 provides more details of the ten clusters. The ten clusters cover 97% of the entire listings.

Seventy-seven percent of the ten clusters' listings connect only to listings in the same cluster. For the remaining listings, Figure 7 presents a histogram of the fraction of neighbors that are in a different cluster. From Figure 7, it can be seen that most of the listings have the majority of their connections in the same cluster.

cluster	color	description	size
1	orange	entire home/apts in San Francisco (excluding the northwestern area)	2566
2	cyan	entire home/apts to the south of South San Francisco and north of San Jose	2100
3	green	private rooms in San Francisco	2093
4	purple	entire home/apts in San Jose	1908
5	magenta	private rooms in San Jose	1629
6	gold	entire home/apts in the northwestern area of San Francisco	1535
7	pink	entire home/apts in Oakland	1390
8	red	private rooms to the south of San Francisco and north of San Jose	1181
9	brown	private rooms in Oakland	590
10	blue	entire home/apts in Daly City and South San Francisco	518

Table 4: Description of the ten clusters visualized in Figure 6. The size of a cluster is the number of listings in the cluster, and the color of a cluster corresponds to its color in Figure 6.

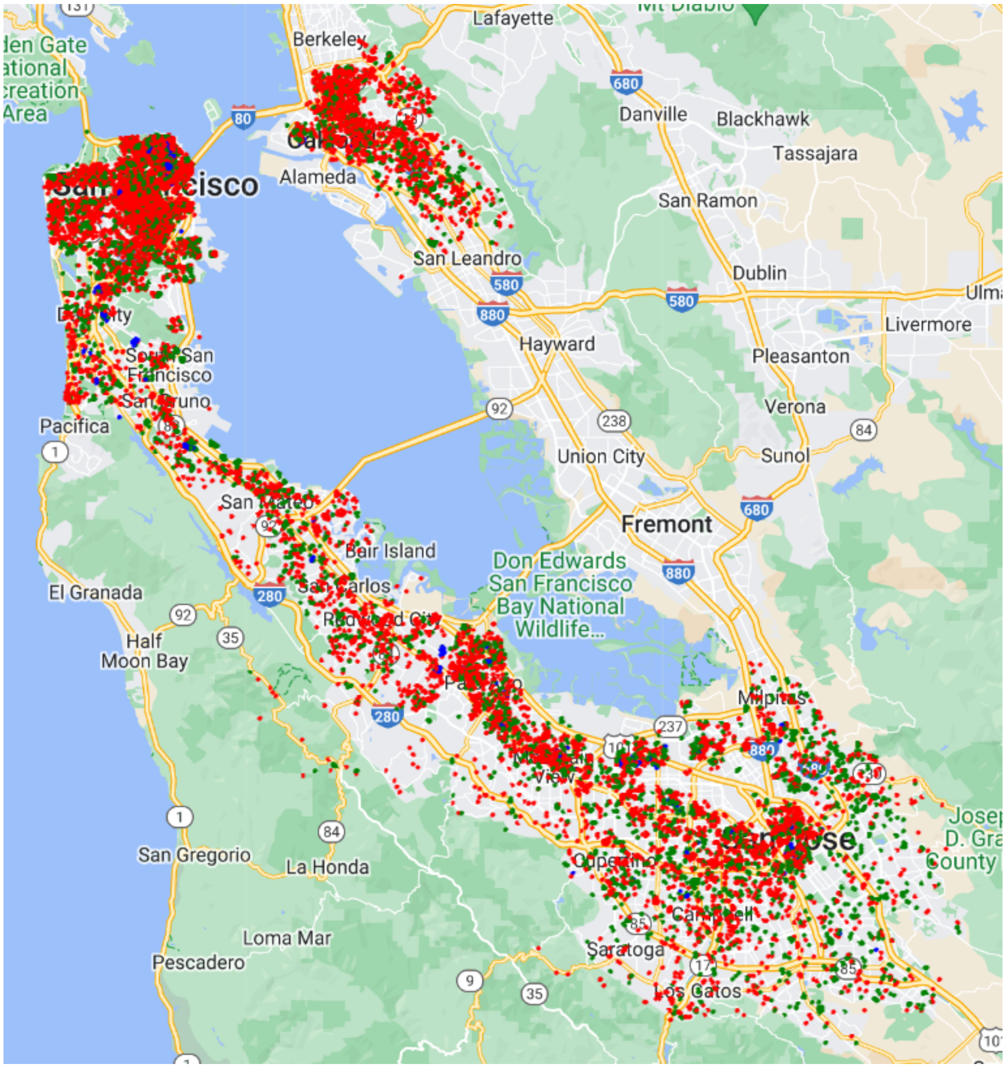


Figure 5: The geographical distributions of Airbnb listings in the Bay area. Red nodes represent listings whose room types are "entire home/apartment" (62.7%), green nodes represent listings whose room types are "private room" (34.5%), and blue nodes represent listings whose room types are either "shared room" or "hotel room" (altogether 2.8%).

D.2 The Optimal Cluster-Based Experiment

An optimal cluster-based experiment for this example is provided in Table 5. The induced correlation structure was reported in Section 5.2. The worst-case value of w_i under the experiment is $y^{\mathrm{OPT}} = [0, 2100, 2093, 0, 1629, 1535, 1390, 0, 0, 0]^{\mathrm{T}}$.

Note that when the marginal assignment probability q equals $\frac{1}{2}$, there always exists an optimal experiment that is symmetric across treatment and control assignments; i.e., letting $P(S)$ denote the probability that clusters in set $S \subseteq [n]$ receive the treatment whereas clusters not in set S receive

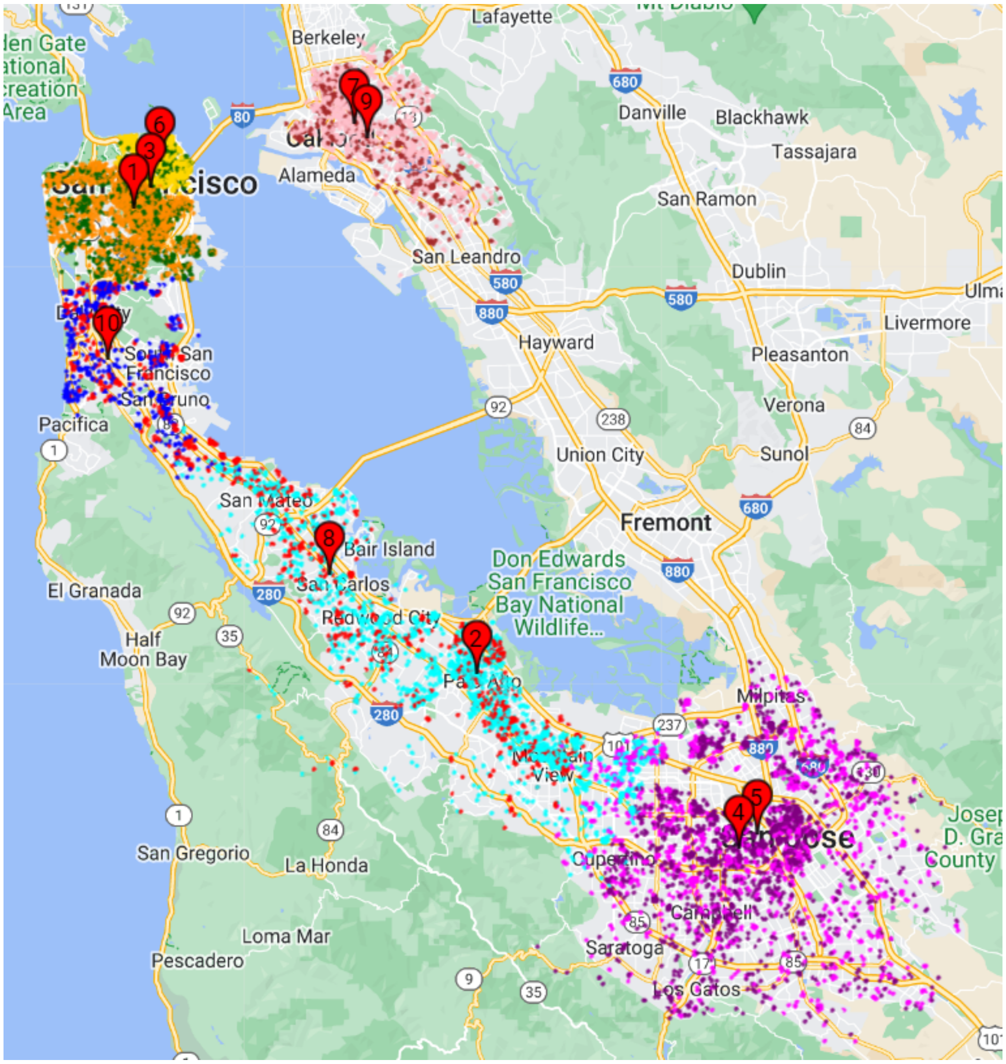


Figure 6: The partition of listings into $n = 10$ clusters, with each color representing a different cluster.

the control, we obtain that $P(S) = P(S^c)$ for any subset $S \subseteq [n]$ of clusters.[15] Such an optimal symmetric experiment still has a complex randomized assignment. Specifically, it randomizes over 88 different possible assignment vectors (where the number of treated clusters varies between 4 and 6), and chooses different probabilities for these vectors without following any clear patterns.

[15]To see this, note that given an optimal experiment, we can always construct another optimal experiment by flipping the treatment and control assignments by Remark 2.1. Then, since the optimal experimental design problem can be formulated as a linear program as in (6), randomizing over the two optimal experiments with equal probability is still optimal, and is symmetric across the treatment and control assignments.

	1	2	3	4	5	6	7	8	9	10	probability
1	√	√	×	×	√	√	×	×	×	×	0.0565
2	√	×	√	√	×	×	√	×	×	×	0.0499
3	√	√	√	×	×	×	×	×	√	√	0.0434
4	×	×	√	√	√	√	×	×	×	√	0.0417
5	×	√	√	×	×	√	√	√	×	×	0.0386
6	√	×	√	×	√	×	×	√	×	√	0.0351
7	×	√	√	×	√	×	×	√	√	×	0.0349
8	√	×	×	×	√	×	√	√	√	×	0.0349
9	√	×	×	√	×	√	×	√	√	×	0.0337
10	×	×	√	√	×	√	×	√	√	√	0.0323
11	×	√	√	√	×	×	×	√	×	×	0.0322
12	×	√	√	×	×	√	√	×	√	×	0.0309
13	√	×	×	×	×	√	√	√	√	√	0.0296
14	×	×	√	×	√	√	√	×	√	√	0.0296
15	×	√	×	√	×	√	√	×	×	√	0.0286
16	√	×	√	×	√	×	√	×	×	√	0.0270
17	×	√	×	√	√	×	√	×	√	√	0.0266
18	√	×	×	×	×	√	√	√	×	√	0.0256
10	√	√	×	√	×	×	×	√	×	√	0.0245
20	×	×	√	√	√	√	×	×	√	×	0.0239
21	√	√	×	√	×	×	×	×	√	√	0.0232
22	×	√	√	×	×	×	√	√	×	√	0.0182
23	√	×	√	√	×	×	×	×	√	×	0.0181
24	√	×	×	√	√	×	√	×	×	×	0.0172
25	×	√	×	√	×	×	√	√	√	√	0.0167
26	×	×	×	√	√	√	×	√	√	√	0.0155
27	√	√	×	×	√	√	×	×	√	×	0.0155
28	×	×	×	√	√	√	√	√	×	√	0.0151
29	×	√	×	√	√	×	√	×	√	×	0.0148
30	×	√	×	√	√	×	√	√	×	×	0.0140
31	√	×	×	√	×	√	√	×	√	×	0.0139
32	×	√	×	×	√	×	√	√	√	√	0.0127
33	√	×	√	×	×	√	×	√	×	×	0.0126
34	×	√	×	×	√	√	×	√	×	√	0.0114
35	√	√	×	×	×	√	×	×	√	√	0.0097
36	×	×	√	×	√	×	√	√	√	×	0.0097
37	√	×	×	√	√	×	×	√	√	×	0.0096
38	×	√	×	√	√	√	×	√	×	×	0.0085
39	×	√	×	√	√	×	×	√	×	√	0.0080
40	×	×	×	√	√	√	√	√	×	×	0.0079
41	×	√	×	√	√	×	√	√	√	×	0.0076
42	√	√	×	×	×	√	√	×	×	√	0.0064
43	×	√	√	×	×	√	√	×	√	√	0.0061
44	×	×	√	√	√	×	√	×	×	×	0.0060
45	×	×	√	√	√	×	√	×	√	√	0.0055
46	√	×	×	×	√	×	√	√	×	√	0.0041
47	√	√	×	√	×	√	×	×	×	×	0.0035
48	√	√	×	×	√	×	×	√	×	√	0.0033
49	×	√	√	×	√	√	×	√	×	×	0.0028
50	√	√	×	√	×	×	√	×	×	×	0.0014
51	√	×	√	×	×	×	√	√	√	×	0.0008
52	√	×	√	×	√	×	√	×	√	×	0.0005
53	√	×	√	×	×	√	×	×	√	×	0.0001

Table 5: The randomized joint assignment of the optimal cluster-based experiment. Each of the 53 rows corresponds to a possible assignment, where $\surd$ denotes treatment and $\times$ denotes control.

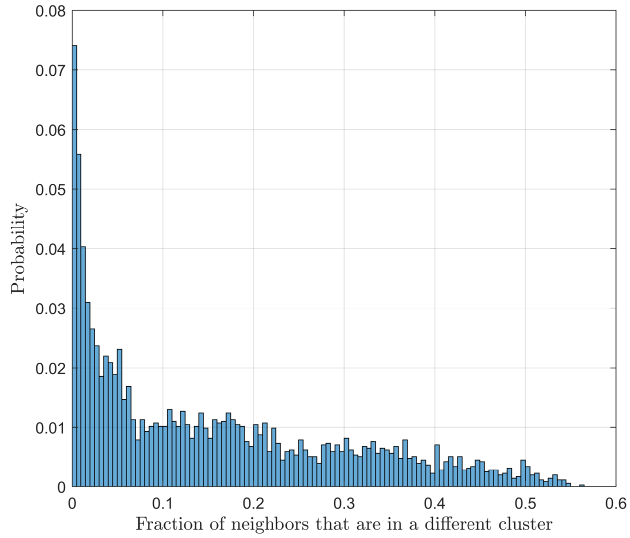


Figure 7: Histogram of the fraction of connections that are in a different cluster, among those listings that ever have a connection to a different cluster (which amount to 23% of the entire listings in the ten clusters).

D.3 The Optimal IBR Experiment and Other Heuristics

The correlation matrix of the optimal IBR experiment (computed by solving a DP) is

$$\Sigma^{\mathrm{DP}} = \begin{pmatrix} 1 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

The worst-case value of w_i for this experiment is $y^{\mathrm{DP}} = [2566, 2100, 0, 0, 1629, 1535, 0, 0, 590, 0]^{\mathrm{T}}$.

The correlation matrix of the `HALF` experiment is

$$\Sigma^{\mathrm{half}} = \frac{10}{9} I - \frac{1}{9}\mathbf{1}\mathbf{1}^{\mathrm{T}};$$

i.e., all the diagonal entries are one and all the off-diagonal entries are $-\frac{1}{9}$. The worst-case value of w_i for this experiment is $y^{\mathrm{half}} = [2566, 2100, 2093, 1908, 0, 0, 0, 0, 0, 0]^{\mathrm{T}}$.

The correlation matrix of the PAIR experiment is

$$\Sigma^{\text{pair}} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

The worst-case value of w_i for this experiment is $y^{\text{pair}} = [2566, 0, 2093, 0, 1629, 0, 1390, 0, 590, 0]^{\mathrm{T}}$.

Finally, the correlation matrix of the IND experiment is the identity matrix $\Sigma^{\text{ind}} = I$. The worst-case value of w_i for this experiment is $y^{\text{ind}} = [2566, 2100, 2093, 1908, 1629, 1535, 1390, 1181, 590, 518]^{\mathrm{T}}$.

E Numerical Example: Facebook Subnetworks of US Universities

In this section, we consider a numerical example based on Facebook subnetworks of one hundred US universities. Specifically, we leverage the data described in Section 2 of Traud et al. (2012), which can be accessed from Rossi and Ahmed (2015). We consider cluster-based experiments over these subnetworks, with the users from each university constituting one cluster. We assume that users from different universities are only loosely connected (in contrast with the dense connection structure within each subnetwork), and that the interference among these subnetworks is negligible. Analogous to Section 5.2, we again focus on the case where the marginal assignment probability q is set to be $\frac{1}{2}$, and we assume that the upper bounds w_{i1} and w_{i0} of the cluster-level treatment and control potential outcomes are both proportional to the size (i.e., number of users) of cluster i; please refer to Traud et al. (2012) for the sizes of these one hundred clusters.

With $n = 100$ clusters, it is computationally prohibitive to obtain the optimal cluster-based experiment. The optimal IBR experiment, on the other hand, is fairly easy to compute by solving the DP in Section 3.2. Specifically, the optimal IBR experiment partitions the clusters into 11 blocks, with the number of clusters in each block being[16]

$$8, 10, 10, 12, 10, 12, 12, 10, 8, 4, 4.$$

[16] Blocks are sorted in decreasing order of size from the largest cluster in each block.

We again consider the three natural experiments in Section 5.2 for comparison. The `HALF` experiment increases the worst-case variance by $\frac{V^{\text{half}}-V^{\text{DP}}}{V^{\text{DP}}}=78.3\%$ relative to our IBR experiment. The `PAIR` experiment increases the worst-case variance by $\frac{V^{\text{pair}}-V^{\text{DP}}}{V^{\text{DP}}}=62.9\%$. The `IND` experiment increases the worst-case variance by $\frac{V^{\text{ind}}-V^{\text{DP}}}{V^{\text{DP}}}=210.0\%$. Thus, the IBR experiment again reduces the worst-case variance considerably when compared to other commonly used heuristic experiments.

E.1 Average-Case Analysis of the Facebook Example

In this section, we conduct an average-case analysis of the Facebook subnetwork example by comparing different experiments' variances under the same potential outcomes (in contrast to their respective worst-case outcomes as considered earlier) with the potential outcomes drawn randomly from a given distribution. We focus on the case where the marginal assignment probability q equals $\frac{1}{2}$, and we compare the optimal IBR experiment and the `HALF`, `PAIR`, and `IND` experiments as described in Section 5.2. With some abuse of notation, we let V^{DP}, V^{half}, V^{pair}, and V^{ind} denote the variances of the Horvitz–Thompson estimator for the optimal IBR experiment, the `HALF` experiment, the `PAIR` experiment, and the `IND` experiment, respectively; these are random variables that depend on the value of the potential outcomes.

We consider the following three cases for the underlying distribution of the cluster-level treatment and control potential outcomes y_{i1} and y_{i0}:

Case 1: Sample $y_{i1}, y_{i0} \sim \text{Unif}[0, w_i]$ for each cluster i, all *i.i.d.*,

Case 2: Sample $y_{i0} \sim \text{Unif}\left[0, \frac{w_i}{2}\right]$, *i.i.d.*, and let $y_{i1} = y_{i0} + 0.2w_i$ for each cluster i,

Case 3: Sample $y_{i0} \sim \text{Unif}\left[0, \frac{w_i}{2}\right]$, *i.i.d.*, and let $y_{i1} = y_{i0} + 0.4w_i$ for each cluster i,

where w_i is the number of users in cluster i. In all three cases, the sampling distribution is independent across clusters; we do so to avoid assuming a specific stylized correlation structure across clusters.

Now, let $\tau_a \triangleq \frac{\tau}{m}$ denote the average treatment effect, where τ is the total market effect and $m = \sum_{i\in[n]} w_i$ is the total number of users. Note that τ_a is approximately zero in Case 1, $\tau_a = 0.2$ in Case 2, and $\tau_a = 0.4$ in Case 3. We further let $\sigma_{\text{DP}} \triangleq \frac{\sqrt{V^{\text{DP}}}}{m}$ denote the standard deviation of the Horvitz–Thompson estimator for τ_a under the optimal IBR experiment. Analogously, we let $\sigma_{\text{half}} \triangleq \frac{\sqrt{V^{\text{half}}}}{m}$, $\sigma_{\text{pair}} \triangleq \frac{\sqrt{V^{\text{pair}}}}{m}$, and $\sigma_{\text{ind}} \triangleq \frac{\sqrt{V^{\text{ind}}}}{m}$ denote the standard deviations of the Horvitz–Thompson estimator for τ_a under the `HALF`, `PAIR`, and `IND` experiments, respectively. In the simulation, we randomly draw the potential outcomes 10^4 times for each case and we present the box plots of the values σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind} and the ratios $\sigma_{\text{half}}/\sigma_{\text{DP}}$ and $\sigma_{\text{pair}}/\sigma_{\text{DP}}$ in Figure 8.

As can be seen from Figure 8, the optimal IBR experiment reduces the variance substantially compared to the HALF and IND experiments in all of the cases and under all realizations of the potential outcomes. The PAIR experiment has a marginally smaller variance on average under the three sampling distributions[17] (see the right column of Figure 8). On the other hand, the optimal IBR experiment attains a smaller worst-case variance, its variance is more concentrated around the median, and hence it is more robust to the unknown potential outcomes. We also highlight that the observation that the PAIR experiment has a smaller variance (on average) than our IBR experiment is substantially an artifact of the choice of the sampling distribution. In particular, the assumption that the potential outcomes are independent across clusters is indeed the source of this observation. When, for example, the potential outcomes are negatively correlated between clusters of similar size, the variance of the PAIR experiment is in general larger than the variance of the optimal IBR experiment.[18] From all three cases in this example, it can be seen that although the optimal IBR experiment is designed with the goal of minimizing the worst-case variance, it maintains the same performance or even reduces the variance on average in comparison to other heuristic experiments.

Finally, we present the box plot of the ratio $\sigma_{\text{DP}}/\tau_a$, which is the standard deviation of the Horvitz–Thompson estimator of the average treatment effect over the true value, for Cases 2 and 3 in Figure 9. In Figure 9, the standard deviation is relatively small compared to the true average treatment effect, and this demonstrates the statistical power of the estimator. We elaborate more on this point in Appendix F.

F Statistical Inference from IBR Experiments

In this section, we discuss a way to construct the confidence interval for estimating the total market effect with an IBR experiment. The Horvitz–Thompson estimator $\hat{\tau}$ is unbiased. Now fix the potential outcomes for treatment and control. When the number of blocks is large, by the central limit theorem and the fact that assignments are independent across blocks, the distribution of the estimator $\hat{\tau}$ is approximately normal. When the number of blocks is small, we assume that $\hat{\tau}$ is approximately normal as well. Hence, an α-level confidence interval for the total market effect can be given by $\left[\hat{\tau} - z_{\alpha/2}\sqrt{\mathbb{V}\text{ar}[\hat{\tau}]}, \hat{\tau} + z_{\alpha/2}\sqrt{\mathbb{V}\text{ar}[\hat{\tau}]}\right]$, where $z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$, with $\Phi(\cdot)$ being the CDF of a standard normal distribution.

The problem, however, is that we are not able to compute the variance $\mathbb{V}\text{ar}[\hat{\tau}]$, as it depends

[17] Specifically, the median of the ratio $\sigma_{\text{pair}}/\sigma_{\text{DP}}$ is 0.98 in Cases 1 and 2 and 0.97 in Case 3.

[18] This can happen when the potential outcomes are correlated through (possibly unobserved) covariates and these covariates are quite different across clusters of similar size.

on *all* the values of potential outcomes, which cannot be observed simultaneously (recall that by Lemma 2.1, $\mathbb{V}\mathrm{ar}[\hat{\tau}] = q(1-q)y^{\mathrm{T}}\Sigma y$ with $y_i = \frac{y_{i1}}{q} + \frac{y_{i0}}{1-q}$ for each cluster $i \in [n]$). We may use the worst-case variance of the IBR experiment as a proxy for $\mathbb{V}\mathrm{ar}[\hat{\tau}]$, but this can be quite loose especially when some of the potential outcomes are observed after the experiment.

Analogously to Section 4.3 of Imai et al. (2009) and Section 4.2 of Bojinov et al. (2020), we consider a conservative estimator for the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$ (which is a variant of a Neymanian conservative variance estimator; see Imbens and Rubin 2015 and Aronow and Middleton 2013). Specifically, for each cluster $i \in [n]$, we let $y_i^{\mathrm{obs}} \triangleq \frac{\sqrt{1-q}}{q} y_{i1} \cdot \mathbb{1}[Z_i = 1] + \frac{\sqrt{q}}{1-q} y_{i0} \cdot \mathbb{1}[Z_i = 0]$ denote the weighted observed outcome for cluster i. We use the following estimator $\hat{\sigma}^2$ for the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$, with

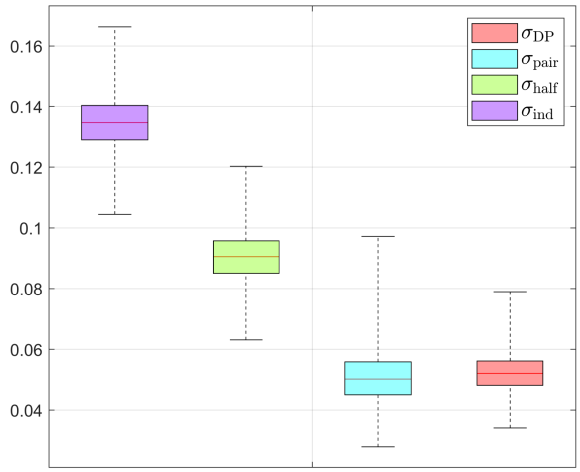
$$\hat{\sigma}^2 \triangleq 2 \sum_{i\in[n]} \left(y_i^{\mathrm{obs}}\right)^2 + \sum_{i\in[n]} \sum_{[n]\ni k\neq i} \sigma_{ik} \left(\left(y_i^{\mathrm{obs}}\right)^2 + \left(y_k^{\mathrm{obs}}\right)^2 \right),$$

where σ_{ik} is the correlation between clusters i and k. The mean of $\hat{\sigma}^2$ provides an upper bound on the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$ because

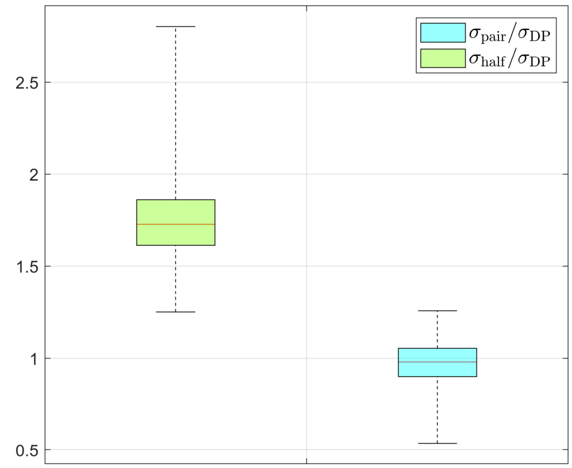
$$\begin{aligned}
\mathbb{E}\left[\hat{\sigma}^2\right] &= q(1-q) \left\{ 2 \sum_{i\in[n]} \left[\left(\frac{y_{i1}}{q}\right)^2 + \left(\frac{y_{i0}}{1-q}\right)^2 \right] + \sum_{i\in[n]} \sum_{k\neq i} \sigma_{ik} \left[\left(\frac{y_{i1}}{q}\right)^2 + \left(\frac{y_{i0}}{1-q}\right)^2 + \left(\frac{y_{k1}}{q}\right)^2 + \left(\frac{y_{k0}}{1-q}\right)^2 \right] \right\} \\
&\geq q(1-q) \left\{ \sum_{i\in[n]} y_i^2 + \sum_{i\in[n]} \sum_{[n]\ni k\neq i} \sigma_{ik} y_i y_k \right\} \\
&= \mathbb{V}\mathrm{ar}[\hat{\tau}],
\end{aligned}$$

where the inequality simply follows from the basic inequality $2xy \leq x^2 + y^2$ and the fact that $y_i = \frac{y_{i1}}{q} + \frac{y_{i0}}{1-q}$.

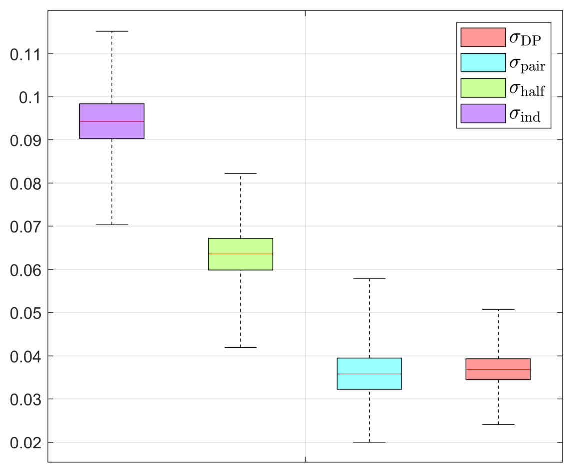
Following Imai et al. (2009) and Bojinov et al. (2020), we suggest using $\left[\hat{\tau} - z_{\alpha/2}\sqrt{\hat{\sigma}^2}, \hat{\tau} + z_{\alpha/2}\sqrt{\hat{\sigma}^2}\right]$ for an α-level confidence interval of the total market effect τ. This is a common heuristic, and we leave its formal analysis for future work.



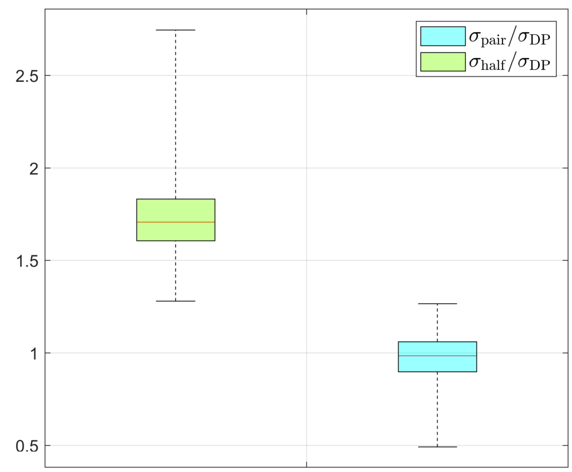
(a) Case 1: σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind}



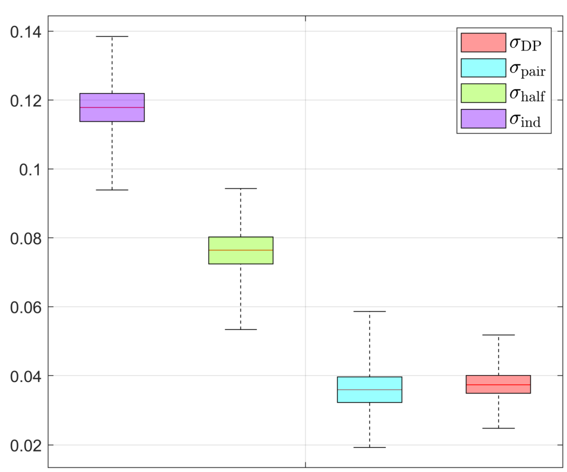
(b) Case 1: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$



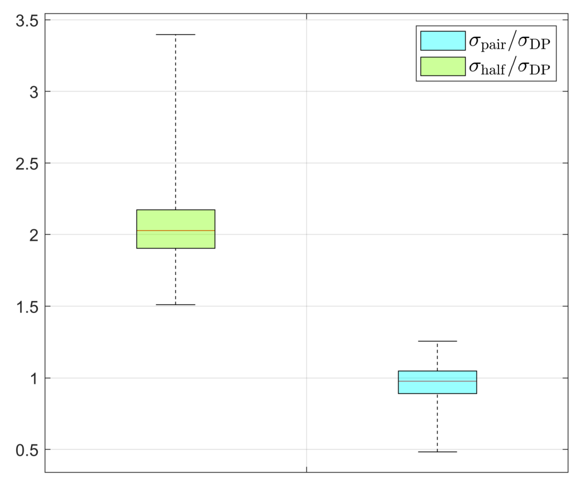
(c) Case 2: σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind}



(d) Case 2: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$



(e) Case 3: σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind}



(f) Case 3: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$

Figure 8: Box plots of the values σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind} (left column) and the ratios $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$ (right column) over 10^4 samples for each case. The interpretation of the box plots is the same as in Figure 3.

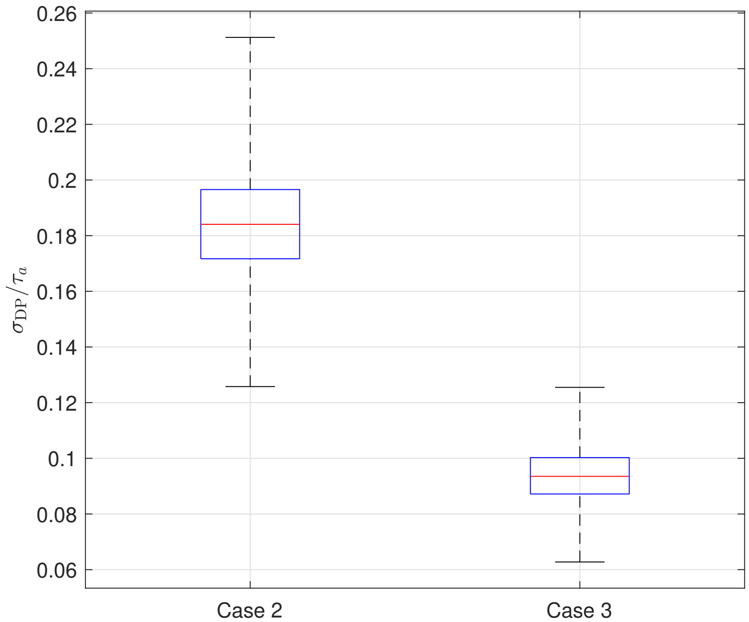


Figure 9: Box plot of the ratio $\sigma_{\rm DP}/\tau_a$ over 10^4 samples for Cases 2 and 3. The interpretation of the box plot is the same as in Figure 3.