Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Near-Optimal A-B Testing

Nikhil Bhat, Vivek F. Farias, Ciamac C. Moallemi, Deeksha Sinha

To cite this article:
Nikhil Bhat, Vivek F. Farias, Ciamac C. Moallemi, Deeksha Sinha (2020) Near-Optimal A-B Testing. Management Science

Published online in Articles in Advance 09 Apr 2020

. https://doi.org/10.1287/mnsc.2019.3424

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–19
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Near-Optimal A-B Testing

Nikhil Bhat,[a] Vivek F. Farias,[b] Ciamac C. Moallemi,[a] Deeksha Sinha[c]

[a] Graduate School of Business, Columbia University, New York, New York 10028; [b] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; [c] Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Contact: nbhat15@gsb.columbia.edu (NB); vivekf@mit.edu, https://orcid.org/0000-0002-5856-9246 (VFF); ciamac@gsb.columbia.edu, https://orcid.org/0000-0002-4489-9260 (CCM); deeksha@mit.edu (DS)

Received: October 22, 2015
Revised: July 19, 2017; September 20, 2018
Accepted: September 22, 2018
Published Online in Articles in Advance: April 9, 2020

https://doi.org/101287/mnsc.2019.3424

Abstract. We consider the problem of A-B testing when the impact of the treatment is marred by a large number of covariates. Randomization can be highly inefficient in such settings, and thus we consider the problem of optimally allocating test subjects to either treatment with a view to maximizing the precision of our estimate of the treatment effect. Our main contribution is a tractable algorithm for this problem in the online setting, where subjects arrive, and must be assigned, sequentially, with covariates drawn from an elliptical distribution with finite second moment. We further characterize the gain in precision afforded by optimized allocations relative to randomized allocations, and show that this gain grows large as the number of covariates grows. Our dynamic optimization framework admits several generalizations that incorporate important operational constraints such as the consideration of selection bias, budgets on allocations, and endogenous stopping times. In a set of numerical experiments, we demonstrate that our method simultaneously offers better statistical efficiency and less selection bias than state-of-the-art competing biased coin designs.

History: Accepted by Noah Gans, stochastic models and simulation.
Funding: This work was supported by the Division of Civil, Mechanical and Manufacturing Innovation. The second author's work was supported in part by United States National Science Foundation (NSF) [CAREER Grant CMMI-1054034]. The third author was supported by NSF [Grant CMMI-1235023].
Supplemental Material: The electronic companion is available at https://doi.org/10.1287/mnsc.2019.3424.

Keywords: decision analysis: inference • dynamic programming • Markov • infinite state • statistics: design of experiments

1. Introduction

The prototypical example of an A-B test is the design of a clinical trial where one must judge the efficacy of a treatment or drug relative to some control. In a different realm, A-B testing today plays an increasingly pivotal role in e-commerce, ranging from the optimization of content and graphics for online advertising, to the design of optimal layouts and product assortments for web pages. E-commerce properties will even use A-B testing as a means of finding the best third-party vendor for a specific service on their website (such as, say, recommendations or enterprise search).

A natural approach to A-B testing is to independently, and with equal probability, assign each subject to either the treatment or the control group. Following such a randomized allocation, the benefit of the treatment relative to the control can be estimated from the outcomes of subjects in the two groups. The notion of a subject here can range from a patient in the clinical trial setting to a web surfer or impression in the e-commerce setting. Similarly, the notion of a treatment can vary from an actual medical treatment in the clinical trial setting to the decision to show a specific ad in the e-commerce setting. Although randomized allocation is simple and can easily be shown to yield unbiased estimates of the treatment effect under a minimal set of assumptions, the efficiency of this procedure (or the sample size needed to get a statistically significant estimate of the treatment effect) can prove onerous in practice. To see, why consider the following challenges:

1. Limited sample size: In the clinical trial setting, the number of subjects is limited for several reasons. As an example, the cost of managing a single subject through a clinical trial is tens of thousands of dollars (see, e.g., Steensma and Kantarjian 2014). In the e-commerce setting, one may need to conduct many thousands of A-B tests in an ongoing fashion. As an example, consider an advertising firm that uses A-B testing on live impressions (i.e., web surfers) to mechanically decide the appropriate messaging, text size, font, color, and so on, for the creative content it generates for an online advertising campaign. In this domain, a reduction in the sample size needed to learn can, due to scale, result in dramatic, continual cost savings.

2. Confounding effects: Running counter to the need for quick inference, the impact of a particular treatment (or design decision) may be marred by a potentially large number of covariates. The presence of these covariates makes the inference of the treatment effect more challenging, since the difference in outcome of the treatment and control groups might be due to a lack of balance in the covariates in the two groups. Although the law of large numbers assures us that a large enough sample size will wash out the impact of this imbalance of covariates, the requisite sample size may grow exceedingly large when the number of covariates is large and/or the treatment effect is small.

3. Small treatment effects: Similar to the covariate imbalance issue, the incremental impact of the treatment under study may be relatively small. This creates a challenge in the measurement of small treatment effects, which, despite their magnitude, many nevertheless be important in settings where the selected treatments will be applied on a sufficiently large scale. More precisely, if one imagined a model where the outcome is additively impacted by the treatment and exogenous noise, we expect the sample size required to discern the treatment from noise to grow quadratically with the ratio of the standard deviation of the exogenous noise to the treatment effect. To (heuristically) see why, observe that if S_n is the sum of n independent, zero mean random variables, each with standard deviation σ, $\theta > 0$ is some constant, and $\Phi(\cdot)$ is the cumulative distribution of the standard normal, then by the central limit theorem, we expect

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \theta\right) \sim 2\Phi\left(\frac{\theta\sqrt{n}}{\sigma}\right).$$

This suggests that, in order to differentiate a treatment effect with magnitude θ from exogenous noise with standard deviation σ, we need on the order of σ^2/θ^2 samples.

4. Operational constraints: As already alluded to, A-B tests can be expensive, either because of an explicit cost related to managing test subjects or the implicit risk of testing a suboptimal treatment. These issues clearly impact the choice of sample size and frequently imply a budget on the number of subjects allocated to the alternative treatment whose efficacy we seek to measure. It is also not unusual to dynamically stop a trial based on one's confidence in the outcome. In clinical trials, one cares about selection bias in addition to efficiency; measures such as selection bias speak to concerns of robustness (to modeling errors or manipulation), or even fairness. Taken together, these operational constraints further complicate an already challenging problem.

Addressing these challenges motivates considering the careful design of such A-B tests. In particular, given a collection of subjects, some of whom must be chosen for treatment, and others assigned to a control, we would like an assignment that balances the distribution of covariates across the two groups. This in turn could conceptually yield an efficient estimate of the treatment effect, the primary concern alluded to earlier.

Given the broad applicability of an efficient A-B test, it is perhaps not surprising that a large body of literature within the statistical theory of the design of experiments has considered this very problem, starting with the nearly century old work of Fisher (1935). Although we defer a review of this substantial literature to Section 1.2, a very popular approach to dealing with the problem of achieving covariate balance is the use of stratification. In this approach, the subjects are divided into a number of groups based on the covariates. In other words, the covariate space is divided into a number of regions and subjects whose covariates lie in a certain region are grouped together. Further, each of the groups is randomly split to be allocated to the treatment or the control. Unfortunately, stratification does not scale gracefully with the number of covariates since the number of groups required in stratification will grow exponentially with the dimension.[1] Another natural idea would be to match subjects with similar covariates, followed by assigning one member of a match to the treatment and the other to the control. Such a design would try to mimic an idealistic scenario in which, for n subjects under the experiment, we have $n/2$ pairs of twins. If the matched subjects are indeed close to each other in the space of covariates, we would have that the distribution of covariates in the treatment and control is close to each other, which would cancel out the effect of these covariates. Although this latter approach does allow us to consider a large number of covariates, the literature only appears to present heuristics motivated by these ideas.

To add a further challenge beyond those already discussed, an additional (and very important) requirement apparent from the applications we have described is that the process of allocating subjects (or impressions) to a particular treatment (or creative) must be made sequentially, in an online or dynamic fashion. Again, there is a literature on dynamic allocation starting with seminal work by Efron (1971) on biased coin designs (BCDs). Although a BCD seeks to balance the number of subjects in the treatment and control groups, there is by now a robust literature on so-called covariate adaptive BCDs. These schemes extend Efron's original proposal so that one cares about balance in not just the number of subjects across the two groups but also in the covariate distribution. Viewed from the perspective of dynamic optimization, all of these heuristics can be seen as myopic schemes that in making

an allocation at a given point in time fail to hedge against the future stream of arriving subjects. In fact, the literature surprisingly does not consider the design of an optimal online allocation of subjects to treatments—or online A-B testing in our parlance—as a principled dynamic optimization problem where dynamic programming techniques for optimal sequential decision making can be applied.

The present paper casts the problem of computing an efficient estimate of the treatment effect in an A-B test as a dynamic optimization problem. Despite this being a high-dimensional control problem, we show that one can efficiently compute near-optimal solutions to this problem when covariates are elliptically distributed. We show that our approach yields Pareto improvements over state-of-the-art alternatives covariate adaptive BCD approaches. As a secondary contribution, we also show that that the important offline variant of the problem also admits an efficient optimal algorithm and tightly characterize the value of optimization in that setting.

1.1. This Paper

Our approach, in a nutshell, is to formulate online A-B testing as a (computationally challenging) dynamic optimization problem and develop approximation and exact algorithms for the same. In particular, the present paper considers the setting where a subject's response is linear in the treatment and covariates. As we discuss later, this is a canonical model and is widely encountered in the literature on experiment design. We consider the problem of maximizing the precision of our estimate of the treatment effect by optimally allocating subjects to either the treatment or control group. We formulate this problem as a dynamic optimization problem and make the following contributions:

1. Offline allocation: In the offline setting, that is, where the allocation can be made after observing all subjects, we show that the problem can be solved efficiently by using as a subroutine a generalization of the MAX-CUT semidefinite program (SDP) relaxation of Goemans and Williamson (1995). Although not our main result, this result shows that the problem of offline A-B testing (which is still valuable in some traditional applications) can surprisingly be solved efficiently. We also characterize the value of optimized allocations relative to randomization in this setting and show that this value grows large as the number of covariates grows.

2. Sequential allocation: In the online setting, which is the algorithmic focal point of our work, our optimization problem is, not surprisingly, a high-dimensional dynamic optimization problem with dimension that grows like the number of covariates. We show how to break the curse of dimensionality here. In particular, we show that the state space of this dynamic optimization problem collapses if covariates come from an elliptical family of distributions (a family that includes, for example, the multivariate Gaussian). This yields an efficient algorithm that is provably optimal in the elliptical distribution setting and that can nonetheless be employed when covariates are not from an elliptical family.

3. A general framework: We show that our dynamic optimization formulation permits the consideration of criteria beyond just the variance of the treatment effect. Specifically, we extend our formulation to a framework that can accommodate the simultaneous minimization of selection bias; the minimization of general separable cost functions of the allocation; endogenous (optimal) stopping criteria (as opposed to a priori fixed sample sizes); and budgets on the sample size for a given treatment, to name just a few applications of the framework.

4. Experimental comparisons: We compare our approach to sequential allocation with a host of so-called covariate adaptive BCD approaches, several of which are considered state of the art. It is typical to measure the performance of such approaches not just in terms of efficiency, but also with respect to the so-called selection bias they induce. Here we show that our approach yields a Pareto improvement over these alternatives. In addition to synthetic data, we run our experiment on real user impression data from Yahoo.com. We show similar Pareto gains despite the fact that the covariates in the real data are categorical.

Thus, our main contribution is providing an algorithm for the challenging problem of sequential A-B testing that can be shown to be near optimal when covariates are drawn from an elliptical family. The algorithm is applicable to a canonical family of treatment models and also applies to the simultaneous optimization of several criteria. Given the vast extant literature on this problem, and the fact that it is nominally high dimensional, it is a pleasant surprise that such an algorithm exists.

1.2. Related Literature

The theory of optimal experiment design (which, in a sense, subsumes the problems we consider here) starts with the seminal work of Fisher (1935). Important textbook expositions of this mature topic include that of Pukelsheim (2006) and Cook et al. (1979), the latter of which discusses the notion of covariate matching as it applies to practice. Although not our primary focus, the offline problem we discuss in this paper is of practical relevance in the social sciences [see Raudenbush et al. (2007) for an application and heuristics]. Kallus (2013) studies an approach to this problem based on linear mixed integer optimization with an application to clinical trials. In a follow-up paper, Bertsimas et al. (2015)

presents a robust optimization framework for the offline problem with an emphasis on allocations of treatments that are robust to the specific form of the model of each subject's response as a function of the treatments and subject covariates (we merely consider linear functions here). The value of optimization has also recently received attention from the economics community. Kasy (2016) discusses several optimization formations that complement those proposed by Bertsimas et al. (2015). Unlike Kallus (2013) and Bertsimas et al. (2015), Kasy (2016) offers no algorithmic approach to solve the problems he proposes (and unfortunately, his problem formulations appear largely intractable). In contrast, we focus on a class of models where the treatment effect is linear in the observed covariates and offer efficient approximation algorithms for the same. Our formulation is closely related to the case of squared loss with a noninformative prior in the verbiage of Kasy (2016). Our offline problem may be viewed as a special case of the problem of D_a-optimal experiment design and fortuitously coincides with an optimality criterion that already enjoys wide acceptance. By virtue of their computational efficiency, our techniques can be brought to bear in settings where the size of the problem can be very large rendering brute-force techniques for optimization (such as those suggested by Kasy 2016) infeasible.

The problem that is of greatest algorithmic interest to us is the online allocation problem, where treatments must be assigned to subjects as they arrive. With regard to this sequential problem, Efron (1971) proposed an allocation strategy, BCD, that sought to balance the number of subjects in each trial while minimizing certain types of selection bias. Now, whereas Efron's BCD seeks only to balance the number of subjects between test and control groups, there is by now a robust literature on so-called covariate adaptive BCDs (CA-BCDs). Such schemes seek balance not just in the number of subjects but also in the covariate distribution between groups. Perhaps the most widely used CA-BCD is the procedure proposed by Pocock and Simon (1975), wherein the authors recommend a bias that depends on a generic cost function of the covariate imbalance between the two groups. Atkinson (1982, 1999) proposed the first CA-BCD whose design is rooted in theory, specifically to the notion of D_a optimality in experiment design; of course, this approach comes at the cost of assuming a treatment effect model. A number of model-based CA-BCD proposals have followed, including Smith's rule (Smith 1984a, b), the Bayesian procedure of Ball et al. (1993), and rule ABCD, proposed by Baldi Antognini and Zagoraiou (2011), to name a few. The so-called minimization approach of Pocock and Simon (1975) (which applies to generic cost functions of covariate imbalance) has also been recently analyzed by Hu and Hu (2012), who prescribe a more refined class of cost functions that lead to asymptotic balance. Alternatives to the CA-BCD procedure have also been proposed recently: Kapelner and Krieger (2014) presents an approach to achieving covariate balance based on ideas from the theory of online matching.

Viewed from the perspective of dynamic optimization, except for the heuristic proposed by Kapelner and Krieger (2014), all of these approaches can be regarded as myopic policies. Such policies only consider the immediate impact of an allocation decision, and do not consider the impact on future decisions. In general, myopic policies will not be optimal. It is worth noting that for all of the aforementioned procedures, the theoretical analysis available, if any, is always in a limiting regime where sample size grows large keeping the number of covariates fixed. Little is understood in finite samples. More generally, Rosenberger and Sverdlov (2008, p. 406) note that "very little is known about the theoretical properties of covariate-adaptive designs." In contrast, we see that our approach yields provably optimal allocations in finite samples for a host of optimality criteria. As we see in our experimental work, this also translates to Pareto improvements over several of the schemes described earlier, even on real data. It is worth noting, however, that such statements of optimality require restrictions on the types of treatment models one can consider, as well as distributional assumptions on the covariates.

1.2.1. Related but Distinct Problems. It is important to distinguish the experiment design problems considered here from bandit problems, particularly those with side information (e.g., Woodroofe 1979, Langford and Zhang 2007) as both classes of problems frequently find application in very related applications. In theory, the experimental design setting is appropriate when an irrevocable decision of what treatment is appropriate must be made (e.g., the number of ads to display with search results), whereas the bandit setting is appropriate in a setting where the decision can be changed over time to optimize the (say) long-run average value of some objective (e.g., maximizing revenues by finding the best audience for a specific campaign). In practice, the choice of which framework to use is frequently complicated by operational considerations. For instance, consider the problem of deciding between two distinct creatives in an advertising campaign. The bandit formulation is elegant and quite natural for this setting (Hauser et al. 2009, Schwartz et al. 2017). Despite this, it is common industry practice to make such decisions using frequent A-B tests.[2] From a methodological perspective, an important difference is that solution methods for bandit problems need to address an exploitation-exploration trade-off between learning the best alternative and

collecting rewards to optimize the objective; there is no such trade-off in our experimental design setting.

Other problems in marketing science are also close in spirit to the A-B testing problem we study. Adaptive conjoint analysis seeks to learn the tastes of an individual (or a group of individuals) by asking a sequence of questions (or presenting a sequence of choices). In an effort to learn accurately with as small a number of questions, Toubia et al. (2003, 2004) propose a dynamic optimization procedure that is in the spirit of the ellipsoid method in convex optimization.

Another closely related class of problems are ranking and selection problems, where the task is to pick the best of a set of alternatives with a budget on samples (for an overview, see Kim and Nelson 2006). In our lexicon, the emphasis in such problems is choosing from multiple (typically, greater than two) treatments in the absence of observable covariates on a sample. Interestingly, recent progress on this class of problems has also heavily employed dynamic optimization techniques (see, e.g., Chick and Gans 2009, Chick and Frazier 2012, Chick et al. 2017).

As a final note, the major emphasis in our work is on A-B testing with a fixed budget on samples. It is interesting to consider A-B tests that can be stopped with continuous monitoring. Doing so can introduce a significant bias toward false discovery. Johari et al. (2017) have recently made exciting progress on this problem.

2. Model

In this section, we describe the model. Given the model assumptions in Section 2.1, our problem is to maximize the precision of our estimate of the treatment effect. In Section 2.2, we pose the two optimization problems that are of interest. One of them is the offline problem where all subjects can be observed before making allocation decisions and the other is the sequential problem where subjects must be allocated without knowing the future arrivals. In Section 2.3, we present a simple upper bound on the precision of any estimate of the treatment effect given an allocation. This allows us to define the notions of efficiency and loss. Section 2.4 concludes with an intuitive interpretation of our optimization problems.

2.1. Setup

We must learn the efficacy of a treatment by observing its effect on n subjects. The kth subject is assigned a treatment $x_k \in \{\pm 1\}$. The kth subject is associated with a covariate vector (i.e., side information or context) $Z_k \in \mathbb{R}^p$. We assume that impact of the treatment on the kth subject is given by:

$$y_k = x_k\theta + Z_k^\top \kappa + \epsilon_k.$$

This assumes a linear dependence of the covariates and treatment decision on the outcome. The treatment effect $\theta \in \mathbb{R}$ and the weights on the covariates $\kappa \in \mathbb{R}^p$ are unknown. Our aim is to estimate θ. The $\{\epsilon_k\}$ are independent and identically distributed (i.i.d.) zero mean random variables with variance σ^2. The key restriction imposed by this model is that the impact of treatment is additive, an assumption that is ubiquitous in all of the related literature on the topic. Further, we assume that there is no endogeneity, that is, the idiosyncratic noise in the model, ϵ_k, is uncorrelated with any of the covariates in Z_k.[3]

Letting $Z \in \mathbb{R}^{n\times p}$ be the matrix whose kth row is $Z_k^\top$, throughout this paper, we will assume the following.

Assumption 1. *The first column of Z is a vector of all ones. Further, Z is full rank and $p \le n-1$.*

The requirement that one of the covariates be a constant ensures that θ is interpreted as a treatment effect, otherwise it could be learned from the assignment of a single treatment. The crucial assumption is that $p \le n-1$, which nonetheless allows for a large number of covariates.[4] In fact, the scenario where $p \sim n$ is particularly relevant. Our problem formulation does not apply to the regime where $p > n$; indeed, a formulation that is relevant to that regime is unclear to us since treatments must be assigned prior to having observed outcomes. For a particular allocation of treatments, x, let us denote by $\hat{\theta}_x$ the least squares estimator for θ.

2.2. Optimization Problem

We are interested in finding an experiment design with minimal variance or, equivalently, maximal precision. A standard calculation yields that the estimator $\hat{\theta}_x$ has precision

$$\mathrm{Prec}\left(\hat{\theta}_x\right) \triangleq \frac{1}{\mathrm{Var}\left(\hat{\theta}_x\right)} = \frac{x^\top P_{Z^\perp} x}{\sigma^2}, \tag{1}$$

where $P_{Z^\perp} \triangleq I - Z(Z^\top Z)^{-1}Z^\top$. Details are presented in the electronic companion to this paper.

We can now immediately state the offline experiment design problem:

$$\text{(P1)} \triangleq \begin{array}{ll}\text{maximize} & x^\top P_{Z^\perp} x \\ \text{subject to} & x \in \{\pm 1\}^n.\end{array}$$

Here, given the collection of covariates Z, we seek to find the allocation x that yields the least squares estimate with maximal precision.

In many real-world applications, the assignments need to be made in a sequential fashion. Subjects arrive one at a time and the assignment must be made without the knowledge of subjects in the future. We formulate this as a dynamic optimization problem. To this end we must now assume the existence of a

measure on the covariate process $\{Z_k\}$. We define a filtration $\{\mathcal{F}_k\}$ by setting, for each time k, $\mathcal{F}_k$ to be the sigma algebra generated by the first k covariates $(Z_1, \ldots, Z_k)$ and the first $k-1$ allocations $(x_1, \ldots, x_{k-1})$. The online experiment design problem is then given by:

$$\begin{aligned} \text{(P2)} \triangleq \text{maximize} \quad & \mathsf{E}\left[x^\top P_{Z^\perp} x\right] \\ \text{subject to} \quad & x \in \{\pm 1\}^n, \\ & x_k \text{ is } \mathcal{F}_k\text{-measurable}, \quad \forall\, 1 \le k \le n, \end{aligned}$$

where the expectation is over the distribution of the covariate process. Here, the objective is to maximized the expected ex post precision.[5]

2.3. Upper Bound, Efficiency, and Loss

The following upper bound on the precision of any unbiased estimator that is a straightforward consequence of the Cramér-Rao bound.

Proposition 1. *If $\epsilon \sim N(0, \sigma^2 I)$, then for any covariate matrix Z and any unbiased estimator $(\hat{\theta}, \hat{\kappa})$, including non-least squares estimators, we have*

$$\mathrm{Prec}\left(\hat{\theta}_x\right) \le \frac{n}{\sigma^2},$$

an upper bound on the optimal value of both problems (P1) and (P2). For non-Gaussian noise ϵ, this upper bound still holds for all least squares estimators.

This proposition, whose proof is provided for completeness in the electronic companion to this paper, shows that the precision of the optimal estimator[6] is $O(n)$. Consider the case when subjects are identical, that is, $p = 1$ and $Z_k = 1$ for all k. It is easy to note that, in this case assuming n is even, the optimal design allocates half of the subjects to either treatment. Further, the precision of such a design is n/σ^2, the optimal achievable precision. For $p > 1$, this precision is less than this value. Thus the presence of covariates only makes the inference challenging.

Motivated by Proposition 1, we define efficiency as the precision of an estimator normalized by the Cramér-Rao upper bound, that is,

$$\mathrm{Eff}\left(\hat{\theta}_x\right) \triangleq \frac{\mathrm{Prec}\left(\hat{\theta}_x\right)}{n/\sigma^2} \le 1,$$

Loss is defined as the suboptimality of an estimator relative to the upper bound measured additively in sample units:

$$\mathrm{Loss}\left(\hat{\theta}_x\right) \triangleq n - \sigma^2 \mathrm{Prec}\left(\hat{\theta}_x\right) \ge 0,$$

so that

$$\mathrm{Prec}\left(\hat{\theta}_x\right) = \frac{n - \mathrm{Loss}\left(\hat{\theta}_x\right)}{\sigma^2}.$$

We consequently see that loss can intuitively be thought of as "the effective number of subjects on whom information is lost due to the imbalance of the design" (Atkinson 2014, p. 147).

2.4. Problem Interpretation

Before moving on to algorithm design, we pause to interpret the offline and online problems presented previously. First we begin with an intuitive interpretation of the objective. Define the imbalance vector in covariate values between the test and control groups, $\overline{\Delta}_n \in \mathbb{R}^p$, according to $\overline{\Delta}_n \triangleq \sum_{k=1}^n x_k Z_k = Z^\top x$. Notice that the empirical second moment matrix for the covariates is given by $\Gamma_n \triangleq Z^\top Z/n$. Then, it is easy to see that the objective of the offline problem (P1) reduces to

$$x^\top P_{Z^\perp} x = x^\top \left(I - Z\left(Z^\top Z\right)^{-1} Z^\top\right) x = n\left(1 - \overline{\Delta}_n^\top \Gamma_n^{-1} \overline{\Delta}_n\right).$$

Therefore, the offline problem (P1) is equivalent to minimizing the square of the weighted Euclidean norm of $\overline{\Delta}_n$,

$$\left\|\overline{\Delta}_n\right\|_{\Gamma_n^{-1}}^2 \triangleq \overline{\Delta}_n^\top \Gamma_n^{-1} \overline{\Delta}_n,$$

whereas (P2) seeks to minimize the expected value of this quantity where the expectation is over the covariate process and our allocations. Put simply, both problems seek to minimize the aggregate imbalance of covariates between the treatment and control groups, measured according to this norm.

As a final point, we note that the measure of imbalance minimized in problems (P1) and (P2) was derived assuming a least squares estimator, and it is worth noting that this choice is not arbitrary. Specifically, note that the Cramér-Rao bound dictates that, provided x and Z are independent of ϵ, and further if ϵ is normally distributed, then for any unbiased estimator of the treatment effect $\tilde{\theta}_x$, we have that

$$\mathrm{Eff}\left(\tilde{\theta}_x\right) \le \mathrm{Eff}\left(\hat{\theta}_x\right),$$

where the right-hand side quantity is the efficiency of the least square estimator. Now both problems (P1) and (P2) seek to find an allocation x to maximize the latter quantity, or its expected value, respectively. Consequently, both problems may be interpreted as seeking an allocation of samples to the test and control group with a view to maximizing the efficiency of our estimate of the treatment effect among all unbiased estimators of the treatment effect.

3. The Offline Optimization Problem

In this section, we consider the offline optimization problem (P1). We show that this combinatorial problem permits a tractable, constant factor approximation using an SDP-based randomized rounding algorithm.

Moreover, in this setting, we can analyze the effect optimization has on the precision of the estimator of the treatment effect, as compared with randomization. To this end, we first obtain the mean precision of the randomized design. Surprisingly, precision is a simple function of n and p and does not depend on the data matrix Z. We show that when $p \sim n$, the randomization is rather inefficient and the precision is $O(1)$. This can be contrasted with the upper bound on precision given by Proposition 1 which is $\Omega(n)$. To conclude the section, we analyze the performance of the optimal allocation assuming a distribution on Z. We show that for any p, the precision of optimal allocation is $\Omega(n)$. Thus concluding that when $p \sim n$, randomization can be arbitrarily bad as compared with the optimal design.

3.1. Approximation Algorithm for (P1)

First, we observe that there is a tractable approximation algorithm to solve the combinatorial optimization problem (P1). In particular, consider the SDP over symmetric positive semidefinite matrices $Y \in \mathbb{R}^{n\times n}$ given by[7]

$$\begin{aligned} \text{(P1-SDP)} \triangleq \quad & \text{maximize} && \operatorname{tr}(P_{Z^\perp} Y) \\ & \text{subject to} && Y_{kk} = 1, \qquad \forall\ 1 \le k \le n, \\ & && Y \succeq 0, \\ & && Y \in \mathbb{R}^{n\times n}. \end{aligned}$$

It is straightforward to see that (P1-SDP) is a relaxation of (P1) in the sense that it achieves higher objective value: given an optimal solution $\hat{x} \in \{\pm 1\}^n$ for (P1), define the symmetric positive definite matrix $\hat{Y} \triangleq \hat{x}\hat{x}^\top \in \mathbb{R}^{n\times n}$. Then, clearly $\hat{Y}$ satisfies the constraints of (P1-SDP). Also, $\operatorname{tr}(P_{Z^\perp}\hat{Y}) = \hat{x}^\top P_{Z^\perp}\hat{x}$, so the objective values for (P1) and (P1-SDP) coincide. Therefore, the optimal objective value of (P1-SDP) must be larger than that of (P1). Moreover, because it is an SDP, (P1-SDP) can be efficiently solved in polynomial time.

Based upon prior work on the MAX-CUT problem (Goemans and Williamson 1995), the following result, due to Nesterov (1997), establishes that (P1-SDP) can be used as the basis of a randomized algorithm to solve (P1) with a constant factor guarantee with respect to the optimal design. The corresponding (randomized) allocation procedure is described in Algorithm 1.

Algorithm 1 (Randomized Allocation Algorithm Based on (P1-SDP))

1. **procedure** SDPAllocation(Z) ▷ Compute an allocation $\tilde{x}$
2. Set $Y^* \succeq 0$ to be an optimal solution of the program (P1-SDP) given the data matrix Z.
3. Set the matrix $V \in \mathbb{R}^{n\times n}$ with columns $v_1, \ldots, v_n \in \mathbb{R}^n$ so that the matrix decomposition $Y^* = V^\top V$ holds.
4. Let $u \in \mathbb{R}^n$ be a vector chosen at random uniformly over the unit sphere.
5. **for** $k \leftarrow 1, n$ **do**
6. $\tilde{x}_k \leftarrow \begin{cases} +1 & \text{if } u^\top v_k \ge 0, \\ -1 & \text{if } u^\top v_k < 0. \end{cases}$
7. **end for**
8. **return** $\tilde{x}$
9. **end procedure**

Theorem 1. *Given a data matrix $Z \in \mathbb{R}^{n\times p}$, set the allocation $\tilde{x} \in \mathbb{R}^n$ according to Algorithm 1. Then,*

$$\mathsf{E}_u\left[\tilde{x}^\top P_{Z^\perp}\tilde{x}\right] \ge \frac{2}{\pi} \max_{x\in\{\pm 1\}^n} x^\top P_{Z^\perp} x,$$

where the expectation is taken over the choice of random vector u in Algorithm 1. In order words, the expected value achieved by the vector $\tilde{x}$ in the offline experiment design problem (P1) is within a constant factor $2/\pi$ of the best possible.

Proof. This theorem is a direct consequence of theorem 3.4.2 of Ben-Tal and Nemirovski (2001). That result states that any quadratic integer optimization problem with objective $x^\top Q x$, such that $x \in \{\pm 1\}^n$, can be approximated within a relative error of $\pi/2$ using the prescribed algorithm, provided Q is positive semidefinite. Since $P_{Z^\perp}$ is positive semidefinite (indeed, it is a projection matrix), the result follows. □

3.2. Optimal Allocations Versus Randomized Allocations

Randomization is the most popular technique used for A-B testing. In what follows, we will compare the performance of randomization to what can be achieved by the optimal offline allocation of (P1).

In its most basic variation, simple randomization partitions the population into two equally sized groups, each assigned a different treatment, where the partition is chosen uniformly at random over all such partitions (for simplicity, we will assume that the population is of even size). Denote by $X_{\text{rand}} \in \{\pm 1\}^n$ the random allocation generated by simple randomization, and denote by $\hat{\theta}_{X_{\text{rand}}}$ the resulting unbiased least squares estimator for θ.

Theorem 2. *If n is even, given a covariate matrix Z, define the expected precision and loss of simple randomization*

$$\begin{aligned} \text{Prec}_{\text{rand}} &\triangleq \mathsf{E}_{X_{\text{rand}}}\left[\text{Prec}\left(\hat{\theta}_{X_{\text{rand}}}\right)\right], \\ \text{Loss}_{\text{rand}} &\triangleq \mathsf{E}_{X_{\text{rand}}}\left[\text{Loss}\left(\hat{\theta}_{X_{\text{rand}}}\right)\right], \end{aligned}$$

where the expectations are taken over the random allocation X_{rand}. Then,

$$\text{Prec}_{\text{rand}} = \frac{n}{\sigma^2}\left(1 - \frac{p-1}{n-1}\right), \quad \text{Loss}_{\text{rand}} = \frac{n}{n-1}(p-1).$$

The proof relies on simple probabilistic arguments and is presented in the electronic companion to this paper. Surprisingly, the precision and loss of the randomized allocation do not depend on the data matrix Z at all, as long as it is full rank and has a constant column.

Comparing with the upper bound of Proposition 1, we notice that in the large sample size regime where $n \to \infty$, simple randomization is asymptotically order optimal in the sense that it achieves precision that grows with order n—the maximum permitted by the upper bound of Proposition 1—when $p \ll n$. This may not be the case when p is close to n, however. For example, if $p = n - 1$, which is the maximum value p can take under Assumption 1, then $\mathrm{Prec}_{\mathrm{rand}} \approx 1/\sigma^2$, which is of constant order. In such a case, the least squares estimator $\hat{\theta}_{X_{\mathrm{rand}}}$ will not asymptotically converge to θ as $n \to \infty$. In general, simple randomization is asymptotically order optimal any time that $p_n = o(n)$ as $n \to \infty$.

Now we consider the performance of the least squares estimator under the optimal design that would be obtained by solving the offline experiment design problem (P1). By construction, the optimal design will clearly have precision that is at least that of the randomized procedure. We would like to understand the magnitude of the possible improvement, however, and to see if it is material. Unlike in the simple randomized case, however, the precision of the optimal design depends on the covariate matrix Z. Moreover, it is difficult to obtain a closed-form expression for this precision as a function of Z.

We can illustrate this with a simple example. Consider the case where $p = n - 1$. The precision of the optimal design is given by

$$\sup_{x \in \{\pm 1\}^n} \frac{x^\top P_{Z^\perp} x}{\sigma^2}.$$

Since $p = n - 1$, the null space of $Z^\top$ is a one-dimensional subspace of $\mathbb{R}^n$. Let $y \in \mathbb{R}^n$ be a nonzero vector such that $Z^\top y = 0$ and $\|y\|_2^2 = 1$. That is, y is a unit vector in the null space of $Z^\top$. It is easy to see that $P_{Z^\perp} = yy^\top$. Thus, the precision of the optimal design is

$$\sup_{x \in \{\pm 1\}^n} \frac{x^\top yy^\top x}{\sigma^2} = \sup_{x \in \{\pm 1\}^n} \frac{(y^\top x)^2}{\sigma^2} = \frac{\|y\|_1^2}{\sigma^2}. \tag{2}$$

Now, consider the following two cases:

1. In case 1, y has only two nonzero components given by $1/\sqrt{2}$ and $-1/\sqrt{2}$. In this case, the optimal precision is $2/\sigma^2$. Thus, in this case, randomization is within a constant factor of optimal.

2. In case 2, y has entries such that $|y_i| = 1/\sqrt{n}$ and $\mathbf{1}^\top y = 0$. In this case, the precision is n/σ^2. Thus, in this case, the optimal design achieves the Cramér-Rao upper bound and the performance is a significant improvement over the randomized design.

The preceding two cases show that depending on the covariate matrix Z (which determines the vector y in the previous discussion), the performance of the optimal design may be a drastic improvement over that of the randomized design. To study the performance of the optimal design, we proceed by making a certain probabilistic assumption on Z. Under this assumption, we will then analyze the distribution of performance of the optimal design. For this purpose, we will assume a distribution on the covariate matrix Z as follows:

Assumption 2. *Given (n, p) with $1 \le p < n$, assume that the covariate matrix $Z \in \mathbb{R}^{n \times p}$ has i.i.d. rows. Further, assume that for each $1 \le k \le n$, the kth row $Z_k \in \mathbb{R}^p$ satisfies $Z_{k,1} = 1$, and that the vector of all components except the first satisfies $Z_{k,2:p} \sim N(0, \Sigma)$, that is, it is distributed according to a multivariate normal distribution with zero mean and covariance matrix $\Sigma \in \mathbb{R}^{p-1 \times p-1}$.*

It is easy to check that, under Assumption 2, the covariate matrix Z will satisfy the full rank condition of Assumption 1 almost surely. Consider a sequence of problems indexed by the sample size n, and where the dimension of the covariates is given by $1 \le p_n < n$. For each n, let $Z^{n,p_n} \in \mathbb{R}^{n \times p_n}$ be the data matrix satisfying Assumption 2. We have the following.

Theorem 3. *Suppose that Assumption 2 holds with $\Sigma = \rho^2 I$. Let x^* be an optimal design obtained by solving (P1) with covariate matrix $Z = Z^{n,p_n}$, and let $\hat{\theta}_{x^*, Z^{n,p_n}}$ be the corresponding least squares estimator of θ. Denote the precision of this estimator by*

$$\mathrm{Prec}_*^{n,p_n} \triangleq \mathrm{Prec}\left(\hat{\theta}_{x^*, Z^{n,p_n}}\right).$$

Then, we have that for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathsf{P}\left(\frac{\mathrm{Prec}_*^{n,p_n}}{n} < \frac{1}{8\pi\sigma^2} - \epsilon\right) = 0,$$

where the probability is measured over the distribution of the covariates.

Theorem 3 states that, with high probability, the optimal offline optimization-based design always yields $\Omega(n)$ precision under Assumption 2. Note that this is true for all possible values of $p_n < n$ with $p_n = n - 1$ being the worst case (the latter fact is established in the proof). In contrast, Theorem 2 establishes that when $p = n - 1$, the precision one expects under randomized allocation is $O(1)$, so the relative improvement from optimization for this value of p is $\Theta(n)$. In other words, if the number of covariates is comparable to the sample size, we might expect dramatic improvements over simple randomization through optimization.

Moreover, whereas the optimal design requires solution of (P1), which may not be tractable, Theorem 1 suggests a tractable approximation that is guaranteed to achieve the same precision as the optimal design up to a constant factor.

The proof of Theorem 3 is presented in the electronic companion to this paper. Here we provide a proof sketch. Let $Z^{n,p} \in \mathbb{R}^{n\times p}$ and $Z^{n,n-1} \in \mathbb{R}^{n\times n-1}$ be two covariate matrices defined on the same probability space (under Assumption 2 with $\Sigma = \rho^2 I$) such that they are identical on the first p columns. We show that $\mathrm{Prec}_*^{n,p} \geq \mathrm{Prec}_*^{n,n-1}$. This establishes that $p = n-1$ corresponds to the worst case precision and allows us to focus on the sequence $\mathrm{Prec}_*^{n,n-1}$. We then analyze the distribution of $Z^{n,n-1}$. We show that $\mathrm{Prec}_*^{n,n-1}$ can be written down as a function of a unit vector in the null space of $(Z^{n,n-1})^\top$, say $y_n \in \mathbb{R}^n$. Further, y_n describes a random one-dimensional subspace of $\mathbb{R}^n$ that is invariant to orthonormal transformations that leave the constant vector unchanged. There is a unique distribution that has this property. We then identify the distribution and compute the precision in closed-form using this distribution. In particular, we show that, as $n \to \infty$,

$$\frac{\mathrm{Prec}_*^{n,n-1}}{n} \to \frac{1}{8\pi\sigma^2},$$

where the convergence is in distribution.

4. Sequential Problem

We now consider the online experiment design problem (P2). Here, decisions must be made sequentially. At each time k, an allocation $x_k \in \{\pm 1\}$ must be made based only on the first k covariates and any prior allocations. In other words, x_k is $\mathcal{F}_k$-measurable.

In this section we show that the optimization problem is tractable. First, we pose a surrogate problem in which the objective of (P2) is simplified. The details of this simplification are provided in Section 4.1. In Section 4.2, we show that the reduction in performance when the surrogate problem is used to device an assignment policy is negligible. Focusing on the surrogate problem, we show that the surrogate problem is a p-dimensional dynamic program in Section 4.3. Surprisingly, if we assume that the data generating distribution for the covariates comes from the so-called elliptical family, then the state space collapses to two dimensions, making the dynamic program tractable. This state space collapse is presented in Section 4.4.

4.1. Formulation and Surrogate Problem

To formulate the sequential problem with an expected value objective, a probabilistic model for covariates is necessary. We will start by making the following assumption.

Assumption 3. *Given (n,p) with $1 \leq p < n$, assume that the covariate matrix $Z \in \mathbb{R}^{n\times p}$ has i.i.d. rows. Further, assume that for each $1 \leq k \leq n$, the kth row $Z_k \in \mathbb{R}^p$ satisfies $Z_{k,1} = 1$, and that the vector $Z_{k,2:p} \in \mathbb{R}^{p-1}$ of all components except the first has zero mean and covariance matrix $\Sigma \in \mathbb{R}^{p-1\times p-1}$.*

Assumption 3 requires that the sequentially arriving covariates are i.i.d. with first and second moments. Assumption 2, by comparison, in addition imposes a Gaussian distribution.

Problem (P2) can be viewed as maximizing the expectation of terminal reward that is given by

$$\begin{aligned} x^\top P_{Z^\perp} x &= x^\top \left(I - Z(Z^\top Z)^{-1} Z^\top\right) x \\ &= n - \frac{1}{n}\left(\sum_{k=1}^n x_k Z_k\right)^\top \Gamma_n^{-1} \left(\sum_{k=1}^n x_k Z_k\right), \end{aligned} \tag{3}$$

where the sample second moment of covariates is given by

$$\Gamma_n \triangleq \frac{1}{n}\sum_{k=1}^n Z_k Z_k^\top.$$

We write this matrix in block form as

$$\Gamma_n = \begin{bmatrix} 1 & M_n^\top \\ M_n & \Sigma_n \end{bmatrix},$$

where

$$\Sigma_n \triangleq \frac{1}{n}\sum_{k=1}^n Z_{k,2:p} Z_{k,2:p}^\top, \quad M_n \triangleq \frac{1}{n}\sum_{k=1}^n Z_{k,2:p}.$$

Here, M_n and Σ_n correspond to sample estimates of the covariate mean and covariance structure, respectively.

We define, for each k, the scalar sample count imbalance $\delta_k \in \mathbb{R}$ and the covariate imbalance vector $\Delta_k \in \mathbb{R}^{p-1}$ by

$$\delta_k \triangleq \sum_{\ell=1}^k x_\ell, \quad \Delta_k \triangleq \sum_{\ell=1}^k x_\ell Z_{\ell,2:p}. \tag{4}$$

The terminal reward (Equation (3)) is equal to

$$x^\top P_{Z^\perp} x = n - \frac{1}{n}\begin{bmatrix} \delta_n & \Delta_n^\top \end{bmatrix} \begin{bmatrix} 1 & M_n^\top \\ M_n & \Sigma_n \end{bmatrix}^{-1} \begin{bmatrix} \delta_n \\ \Delta_n \end{bmatrix}.$$

Problem (P2) is then equivalent to

$$\begin{aligned} \text{(P3)} \triangleq \text{minimize} \quad & \mathsf{E}\left[\begin{bmatrix} \delta_n & \Delta_n^\top \end{bmatrix} \begin{bmatrix} 1 & M_n^\top \\ M_n & \Sigma_n \end{bmatrix}^{-1} \begin{bmatrix} \delta_n \\ \Delta_n \end{bmatrix}\right] \\ \text{subject to} \quad & x \in \{\pm 1\}^n, \\ & x_k \text{ is } \mathcal{F}_k\text{-measurable}, \quad \forall\, 1 \leq k \leq n. \end{aligned}$$

Observe that the objective of (P3) corresponds to n times the loss of the estimator.

As $n \to \infty$, by the strong law of large numbers (under mild additional technical assumptions), $\Sigma_n \to \Sigma$ and $M_n \to 0$ almost surely. Motivated by this fact, in developing an efficient algorithm for (P3), our first move will be to consider a surrogate problem that replaces the sample covariance matrix Σ_n with the

exact covariance matrix Σ and sets the sample mean M_n to the exact mean 0:

$$\begin{aligned}(\text{P3}') \triangleq \ &\text{minimize} && \mathsf{E}\left[\delta_n^2 + \|\Delta_n\|_{\Sigma^{-1}}^2\right] \\ &\text{subject to} && x \in \{\pm 1\}^n, \\ & && x_k \text{ is } \mathcal{F}_k\text{-measurable}, \quad \forall\ 1 \le k \le n.\end{aligned}$$

Here, given an arbitrary covariance matrix $\hat{\Sigma} \in \mathbb{R}^{p-1\times p-1}$, we find it convenient to introduce the norm $\|\cdot\|_{\hat{\Sigma}^{-1}}$ on $\mathbb{R}^{p-1}$ defined by $\|z\|_{\hat{\Sigma}^{-1}} \triangleq (z^\top \hat{\Sigma}^{-1} z)^{1/2}$. In the present context, this norm is typically referred to as a Mahalanobis distance.

The roles of the sample count imbalance δ_n and the covariate imbalance vector Δ_n in the surrogate problem (P3′) are intuitive: requiring δ_n to be small balances the number of assignments between the two treatments (the focus of the so-called BCDs). Requiring the same of Δ_n will tend to balance covariates—when Δ_n is small, the empirical moments of the covariates across the two treatments are close. As discussed in the introduction, heuristics developed in the literature on the design of optimal trials tend to be driven by precisely these two forces.

For the rest of this section, we will focus on the surrogate problem. We want to first justify the use of the surrogate objective. We do this by providing an approximation guarantee in Section 4.2. We then turn our attention on how to solve the surrogate problem via dynamic programming in the subsequent sections.

4.2. Approximation Guarantee for the Surrogate Problem

First, we show that the policy obtained by solving (P3′) is near optimal. Denote by $\hat{\mu}$ the measure over the sequence x_k induced by an optimal solution for the surrogate control problem (P3′), and let μ^* denote the measure induced by an optimal policy for our original dynamic optimization problem (P3). Now, δ_n and Δ_n are random variables given an allocation policy. Given an allocation policy μ, define

$$D_\mu^{n,p} \triangleq \mathsf{E}_\mu\left[\begin{bmatrix}\delta_n & \Delta_n^\top\end{bmatrix}\Gamma_n^{-1}\begin{bmatrix}\delta_n \\ \Delta_n\end{bmatrix}\right]$$

to be the objective value of (P3) under the allocation policy μ with sample size n and covariate dimension p. The following result is demonstrated, without loss of generality, under the assumption that Σ is the identity (otherwise, we simply consider setting $Z_{k,2:p}$ to $\Sigma^{-1/2}Z_{k,2:p}$).

Theorem 4. *Suppose that Assumption 2 holds with $\Sigma = I$ and let $\epsilon > 0$ be any positive real number. Consider a sequence of problems indexed by the sample size n, where the dimension of the covariates is given by $1 \le p_n < n$ and $\gamma_n > 0$ are real numbers such that, for n sufficiently large, $n \ge L\max(p_n, l\log 2/\gamma_n)/\epsilon^2$. Then, as $n \to \infty$*

$$D_{\hat{\mu}}^{n,p_n} \le \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 D_{\mu^*}^{n,p_n} + \gamma_n n^2 + \gamma_n n^2 p_n + O\left(\sqrt{\frac{n}{p_n - 1}}\right).$$

Here, L and l are universal constants. In particular, selecting $\gamma_n \propto 1/n^4$ yields

$$D_{\hat{\mu}}^{n,p_n} \le \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 D_{\mu^*}^{n,p_n} + O\left(\sqrt{\frac{n}{p_n - 1}}\right). \tag{5}$$

This result relies on the use of nonasymptotic guarantees on the spectra of random matrices with sub-Gaussian entries and can be found in the electronic companion to this paper.

The preceding result bounds the objective of the problem (P3) when (P3′) is used to devise an allocation policy. However, we are interested in the objective of the problem (P2), which is the precision or inverse variance of the design corresponding to the policy used. In particular, denote by $\text{Prec}_\mu^{n,p}$ the expected precision of the estimator when allocations are made with a policy μ, for a problem with sample size n and covariate dimension p, that is,

$$\text{Prec}_\mu^{n,p} = \frac{\mathsf{E}_\mu[x^\top P_{Z^\perp} x]}{\sigma^2} = \frac{n - D_\mu^{n,p}/n}{\sigma^2}. \tag{6}$$

Then, we have the following.

Corollary 1. *Suppose that Assumption 2 holds with $\Sigma = I$. Consider a sequence of problems indexed by the sample size n, where the dimension of the covariates is given by $1 \le p_n < n$, and a fixed positive real number $\epsilon > 0$ such that*

$$\epsilon > \sqrt{L \limsup_{n\to\infty} p_n/n},$$

for a universal constant L. Then, as $n \to \infty$,

$$\frac{\text{Prec}_{\hat{\mu}}^{n,p_n}}{\text{Prec}_{\mu^*}^{n,p_n}} \ge 1 - \frac{4\epsilon^3}{(L-\epsilon^2)(1-\epsilon^2)} + o(1).$$

Corollary 1 gives the multiplicative loss in the precision by using an allocation derived from the surrogate problem (P3′) . The multiplicative loss depends on the ratio p/n, which is captured in the choice of ϵ. For small values of ϵ, the ratio of precision obtained by solving (P3′) and (P2) approaches 1. Note that this result holds in an asymptotic regime where p and n both increase to infinity, as long as p/n remains small.

Proof of Corollary 1. Consider Equation (5) in Theorem 4. This holds when

$$n \geq \frac{L \max\left(p_n, l \log 2/\gamma_n\right)}{\epsilon^2}$$

with $\gamma_n = b/n^4$ for some constant b. Equivalently,

$$n \geq \frac{L \max(p_n, 4l \log n + 2l \log b)}{\epsilon^2}.$$

For n sufficiently large, clearly the constraint that $n \geq L(4l \log n + 2l \log b)/\epsilon^2$ will be satisfied. Therefore, combined with the lower bound hypothesized for ϵ, Equation (5) holds as $n \to \infty$.

Using Equation (6),

$$\begin{aligned}
&\mathrm{Prec}_{\mu^*}^{n,p_n} - \mathrm{Prec}_{\hat{\mu}}^{n,p_n} \\
&= \frac{D_{\hat{\mu}}^{n,p_n} - D_{\mu^*}^{n,p_n}}{n\sigma^2} \\
&\leq \frac{\frac{(1+\epsilon)^2}{(1-\epsilon)^2} D_{\mu^*}^{n,p_n} - D_{\mu^*} + O\left(\sqrt{\frac{n}{p_n - 1}}\right)}{n\sigma^2} \\
&= \frac{4\epsilon D_{\mu^*}^{n,p_n}}{n\sigma^2(1-\epsilon)^2} + o(1) \\
&= \frac{4\epsilon}{(1-\epsilon)^2}\left(\frac{n}{\sigma^2} - \mathrm{Prec}_{\mu^*}^{n,p_n}\right) + o(1). \qquad (7)
\end{aligned}$$

The first inequality follows from Theorem 4 and the last equality from Equation (6).

Let $\mathrm{Prec}_{\mathrm{rand}}^{n,p_n}$ denote precision of the randomized policy. Using Theorem 2 and the optimality of μ^*, we have that

$$\begin{aligned}
\frac{n}{\sigma^2} - \mathrm{Prec}_{\mu^*}^{n,p_n} \leq \frac{n}{\sigma^2} - \mathrm{Prec}_{\mathrm{rand}}^{n,p_n} &= \frac{n}{\sigma^2}\frac{p_n - 1}{n-1} \\
&\leq \frac{n}{\sigma^2}\frac{p_n}{n} \leq \frac{\epsilon^2 n}{L\sigma^2}, \qquad (8)
\end{aligned}$$

where the last inequality uses the fact that, by hypothesis, $p_n/n \leq \epsilon^2/L$. Substituting this into Equation (7) we get that

$$\mathrm{Prec}_{\mu^*}^{n,p_n} - \mathrm{Prec}_{\hat{\mu}}^{n,p_n} \leq \frac{4\epsilon^3 n}{(1-\epsilon)^2 L\sigma^2} + o(1).$$

Now, using Equation (8), we get that,

$$\mathrm{Prec}_{\mu^*}^{n,p_n} \geq \frac{n}{\sigma^2}\left(1 - \frac{\epsilon^2}{L}\right).$$

Thus, we have that

$$\begin{aligned}
1 - \frac{\mathrm{Prec}_{\hat{\mu}}^{n,p_n}}{\mathrm{Prec}_{\mu^*}^{n,p_n}} &\leq \frac{4\epsilon n}{\mathrm{Prec}_{\mu^*}^{n,p_n}(1-\epsilon)^2 L\sigma^2} + o(1) \\
&\leq \frac{4\epsilon^3}{(L-\epsilon^2)(1-\epsilon^2)} + o(1).
\end{aligned}$$

This yields the result. □

4.3. Dynamic Programming Decomposition

It is not difficult to see that (P3′) is a terminal cost dynamic program with state $(\delta_{k-1}, \Delta_{k-1}) \in \mathbb{R}^p$ at each time k. The pair (δ_k, Δ_k) can be interpreted as the postdecision state of the dynamic decision problem immediately after the kth allocation. In other words, given the past arrival sequence and actions, (δ_k, Δ_k) summarizes the impact of this past on the future objective. This is formally stated in the following proposition.

Proposition 2. *Suppose that Assumption 3 holds. For each $1 \leq k \leq n$, define the function $Q_k: \mathbb{R} \times \mathbb{R}^{p-1} \to \mathbb{R}$ by the Bellman equation*

$$\begin{aligned}
&Q_k(\delta_k, \Delta_k) \\
&\triangleq \begin{cases} \delta_n^2 + \|\Delta_n\|_{\Sigma^{-1}}^2, & \text{if } k = n, \\ \mathsf{E}\left[\min\limits_{u \in \{\pm 1\}} Q_{k+1}\left(\delta_k + u, \Delta_k + uZ_{k+1,2:p}\right)\right], & \text{if } 1 \leq k < n. \end{cases} \qquad (9)
\end{aligned}$$

Then,

1. At each time k, the optimal continuation cost for the dynamic program (P3′) is given by $Q_k(\delta_k, \Delta_k)$. In other words, this is the expected terminal cost, given the covariates observed and the allocations made up to and including time k, assuming optimal decisions are made at all future times.

2. Suppose the allocation x_k^* at each time k is made according to

$$x_k^* \in \arg\min_{u \in \{\pm 1\}} Q_k\left(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2:p}\right).$$

Then, the sequence of allocations x^* is optimal for the online experiment design problem (P3′).

Proposition 2, whose proof is presented in the electronic companion to this paper, suggests a standard dynamic programming line of attack for the surrogate problem (P3′): optimal continuation cost functions $\{Q_k\}_{1 \leq k \leq n}$ can be computed via backward induction, and these can then be applied to determine an optimal policy. However, the dimension of this dynamic program is given by the number of covariates p. In general, the computational effort required by this approach will be exponential in p—this is the so-called curse of dimensionality. Thus, outside of very small numbers of covariates, say, $p \leq 3$, the standard dynamic programming approach is intractable. However, as we will now see, that the surrogate problem surprisingly admits an alternative, low dimensional dynamic programming representation.

4.4. State Space Collapse

Proposition 2 yields a dynamic programming approach for the surrogate problem (P3′) that is intractable for all but very small values of p. What is

remarkable, however, is that if the covariate data are assumed to have an elliptical distribution, then (P3′) can be solved via a tractable two-dimensional dynamic program. We first present the technical definition.

Definition 1. A random variable X taking values in $\mathbb{R}^m$ has an elliptical distribution if the characteristic function $\varphi\colon \mathbb{C}^m \to \mathbb{C}$ has the form

$$\varphi(t) \triangleq \mathsf{E}\left[\exp(it^\top X)\right] = \exp\left(i\mu^\top t\right)\Psi\left(t^\top \Sigma t\right),$$

for all $t \in \mathbb{C}^m$, given some $\mu \in \mathbb{R}^m$, $\Sigma \in \mathbb{R}^{m\times m}$, and a characteristic function $\Psi\colon \mathbb{C} \to \mathbb{C}$.

Elliptical distributions, studied extensively, for example, by Cambanis et al. (1981), are a generalization of the multivariate Gaussian distribution. The name derives from the fact that if an elliptical distribution has a density, then the contours of the density are ellipsoids in $\mathbb{R}^m$ parameterized by μ and Σ. A useful standard result for us (see, e.g., Cambanis et al. 1981) is that these distributions can be generated by independently generating the direction and the length of the deviation (in $\|\cdot\|_{\Sigma^{-1}}$-norm) from the center μ.

Proposition 3. *If X has an elliptical distribution with parameters μ, Σ, and Ψ, then there exists a nonnegative random variable R such that,*

$$X \overset{d}{=} \mu + R\Sigma^{1/2}U,$$

where U is distributed uniformly on the unit sphere $\{x \in \mathbb{R}^{p-1} \mid \|x\|_2^2 = 1\}$ and U and R are independent.

Thus, any elliptical distribution can be identified with a vector $\mu \in \mathbb{R}^m$, a positive semidefinite matrix $\Sigma \in \mathbb{R}^{m\times m}$, and random variable R taking values on the nonnegative real line. We denote such a distribution by $\mathrm{Ell}(\mu, \Sigma, R)$. It can be shown that if $R^2 \sim \chi_m^2$ is a chi-squared distribution with m degrees of freedom, then $\mathrm{Ell}(\mu, \Sigma, R)$ is a Gaussian distribution with mean μ and covariance Σ. Well-known distributions such as the multivariate t-distribution, Cauchy distribution, and logistic distribution also fall in the elliptical family.

We state the assumption needed for the state space collapse.

Assumption 4. *Given (n, p) with $1 \le p < n$, assume that the covariate matrix $Z \in \mathbb{R}^{n\times p}$ has i.i.d. rows. Further, assume that for each $1 \le k \le n$, the kth row $Z_k \in \mathbb{R}^p$ satisfies $Z_{k,1} = 1$, and that the vector $Z_{k,2:p} \in \mathbb{R}^{p-1}$ of all components except the first is distributed according to $\mathrm{Ell}(0, \Sigma, R)$, where it is assumed that the random variable R has a finite second moment, and further that, without loss of generality,*[8] *$\mathsf{E}[R^2] = p - 1$.*

The following theorem shows how the p-dimensional dynamic program is reduced to a two-dimensional one with Assumption 4.

Theorem 5. *Suppose that Assumption 4 holds. For each $1 \le k \le n$, define the function $q_k\colon \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ according to*

$$q_k(m, \lambda) \triangleq \begin{cases} m^2 + \lambda, & \text{if } k = n, \\ \mathsf{E}\left[\min\limits_{u\in\{\pm 1\}} q_{k+1}\left(m + u, \lambda + 2uRU_1\sqrt{\lambda} + R^2\right)\right], & \text{if } 1 \le k < n. \end{cases} \tag{10}$$

Here, when $k < n$, the expectation is taken over independent random variables U and R that are the random variables in the stochastic decomposition of $Z_{1,2:p}$ from Assumption 4. Then,

1. *At each time k, the optimal continuation cost for the dynamic program (P3′) is given by*

$$Q_k(\delta_k, \Delta_k) = q_k\left(\delta_k, \|\Delta_k\|_{\Sigma^{-1}}^2\right).$$

In other words, this is the expected terminal cost, given then covariates observed and the allocations made up to and including time k, assuming optimal decisions are made at all future times.

2. *Suppose the allocation x_k^* at each time k is made according to*

$$x_k^* \in \arg\min_{u\in\{\pm 1\}} q_k\left(\delta_{k-1} + u, \left\|\Delta_{k-1} + uZ_{k,2:p}\right\|_{\Sigma^{-1}}^2\right). \tag{11}$$

Then, the sequence of allocations x^* is optimal for the online experiment design problem (P3′).

For the case of Gaussian distribution, the recursion (Equation (10)) for solving the dynamic program (DP) can be simplified according to the following corollary.

Corollary 2. *If Assumption 2 holds, then, for $1 \le k \le n$, the functions $q_k^{\mathrm{gauss}}\colon \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ are given by*

$$q_k^{\mathrm{gauss}}(m, \lambda) \triangleq \begin{cases} m^2 + \lambda, & \text{if } k = n, \\ \mathsf{E}\left[\min\limits_{u\in\{\pm 1\}} q_{k+1}^{\mathrm{gauss}}\left(m + u, \left(\sqrt{\lambda} + u\eta\right)^2 + \xi\right)\right], & \text{if } 1 \le k < n. \end{cases} \tag{12}$$

Here, when $k < n$, the expectation is taken over independent random variables $(\eta, \xi) \in \mathbb{R}^2$, where $\eta \sim N(0, 1)$ is a standard normal random variable, and $\xi \sim \chi_{p-2}^2$ is chi-squared random variable with $p - 2$ degrees of freedom.[9]

We provide the proofs for Theorem 5 and Corollary 2 in the electronic companion to this paper. We make the following observations:

1. A key point is that, unlike the standard dynamic programming decomposition of Proposition 2, Theorem 5 provides a tractable way to solve the surrogate problem (P3′), independent of the covariate dimension p.

This is because the recursion (Equation (10)) yields a two-dimensional dynamic program. One of the state variables of this program, m, is discrete, taking values on the integers from $-n$ to n. Further, one can show that, with high probability, the second state variable λ is $O(n^2)$, thereby allowing us to discretize the state space on a two-dimensional mesh. The functions $\{q_k\}$ can be numerically evaluated on this grid via backward induction. Note that since the expectation in Equation (10) is over a two-dimensional random variable, it can be computed via numerical integration. Further details of this procedure are given in Section 6.

2. Moreover, the functions $\{q_k\}$ do not directly depend on the matrix Σ at all and only indirectly depend on time horizon n through the remaining time $k - n$. In fact, they only depend on the covariate dimension p. For example, in the Gaussian case, this means that if these functions are computed offline, they can subsequently be applied to all p-dimensional problem with a Gaussian data distribution.

3. Finally, the algorithm assumes that the covariance matrix Σ is known. This is needed to compute the $\|\cdot\|_{\Sigma^{-1}}$-norm of Δ_k. In practice, Σ may not be known, and may need to be estimated from data. However, observe that Σ depends only on the distribution of covariates across the subject population, not on the outcome of experiments. In the applications we have in mind, there is typically a wealth of information about this population known in advance of the experimental trials. Hence, Σ can be estimated offline even if the number of covariates p is large and the number of experimental subjects n is small.

For example, in an online advertising setting, and advertiser may want to compare two creatives using A-B testing with a limited number of experimental subjects. In advance of any experiments, the advertiser can use historical data from other trials or market surveys over the same population of subjects to estimate Σ.

5. Variations of the Sequential Problem: A Dynamic Programming Framework

The vanilla formulation of the sequential problem (P2) described in Section 2.2 solely optimizes statistical efficiency. In reality, a complete framework must allow the designer to model a number of additional constraints relevant to practical implementation, including budgets on allocations to the treatment arm; controlling selection bias in addition to maximizing efficiency; optimally stopping an experiment if efficiency objectives are met; and so forth. We will establish that the solution approach described in Section 4 applies to a substantially more general class of problem than the vanilla problem (P2).

To setup this dynamic programming framework, we introduce a few new concepts:

- We will think of the allocation at time $1 \le k \le n$ as a bias $v_k \in [0,1]$. Our optimization algorithm will yield the optimal bias at any given point in time, and then we pick an allocation by flipping a coin with this bias, that is, setting

$$x_k = \begin{cases} +1 & \text{with probability } v_k, \\ -1 & \text{with probability } 1 - v_k. \end{cases} \tag{13}$$

This is the same decision space as in a BCD.

- We are given convex stage wise costs, $c\colon [0,1] \to \mathbb{R}$, that are a function of bias. This can capture for instance, the cost of a sample unit; the extent of non-randomness in a given choice of bias, and so on.
- The set of permitted bias v_k at any stage $1 \le k \le n$ can be constrained to an arbitrary convex set that is itself a function of the state at that time, $\mathcal{V}_k(\delta_{k-1}, \|\Delta_{k-1}\|^2_{\Sigma^{-1}}) \subset [0,1]$.
- Instead of a fixed time horizon n, we allow the experiment to be stopped early according to a stopping time $1 \le \tau \le n$. As we discuss later, this allows us to model optimal early stopping based, for instance, on estimating the treatment effect with a desired precision.

Given these concepts, and an arbitrary parameter $\gamma \ge 0$, consider the following generalization of the problem (P3′):

$$\begin{aligned} \text{(P3′′)} \triangleq \text{minimize} \quad & \mathsf{E}\left[\delta_\tau^2 + \|\Delta_\tau\|^2_{\Sigma^{-1}} + \gamma \sum_{k=1}^{\tau} c(v_k)\right] \\ \text{subject to} \quad & v_k \in \mathcal{V}_k\left(\delta_{k-1}, \|\Delta_{k-1}\|^2_{\Sigma^{-1}}\right), \\ & \qquad \forall\ 1 \le k \le n, \\ & v_k \text{ is } \mathcal{F}_k\text{-measurable}, \\ & \qquad \forall\ 1 \le k \le n. \end{aligned}$$

Following the same arguments as in Section 4.4, (P3′′) can be solved according to optimal continuation costs given by the two-dimensional Bellman recursion[10]:

$$q_k(m,\lambda) \triangleq \begin{cases} m^2 + \lambda, & \text{if } k = \tau, \\ \mathsf{E}\Big[\min\limits_{v \in \mathcal{V}_{k+1}(m,\lambda)} \gamma c(v) \\ \quad + v q_{k+1}\left(m+1, \lambda + 2RU_1\sqrt{\lambda} + R^2\right) \\ \quad + (1-v) q_{k+1}\left(m-1, \lambda - 2RU_1\sqrt{\lambda} + R^2\right)\Big], \\ & \text{if } 1 \le k < \tau, \end{cases} \tag{14}$$

for each time k. Given the optimal continuation costs, an optimal decision v_k at each time k can be computed according to

$$v_k^* \in \underset{v \in \mathcal{V}_k\left(\delta_{k-1}, \|\Delta_{k-1}\|^2_{\Sigma^{-1}}\right)}{\arg\min} \gamma c(v) + v q_k\left(\delta_{k-1}+1, \left\|\Delta_{k-1}+Z_{k,2:p}\right\|^2_{\Sigma^{-1}}\right) + (1-v) q_k\left(\delta_{k-1}-1, \left\|\Delta_{k-1}-Z_{k,2:p}\right\|^2_{\Sigma^{-1}}\right), \quad (15)$$

In the following, we illustrate how (P3′′) addresses several practical variations of the sequential allocation problem.

5.1. Selection Bias

An important consideration that has emerged in the literature on A-B testing is managing so-called selection bias. Following Blackwell and Hodges (1957), one commonly defines the selection bias of an allocation over n time steps as $\frac{2}{n}\sum_{k=1}^{n}|v_k - 1/2|$. Notice that perfect randomization has zero selection bias, whereas a fully deterministic procedure (where v_k is either 0 or 1) has the highest bias possible, one.

It is frequently important to balance this bias against efficiency (or, equivalently, loss). In particular, we want a Pareto optimal solution across the two criteria. Atkinson (2014) compares a multitude of state-of-the-art BCD procedures and calls a procedure admissible if it is not Pareto dominated by some other procedure. He finds that none of the heuristics he examines can be ruled out implying that none of these heuristics are Pareto optimal. But by varying $\gamma \geq 0$ in (P3′′), we can generate a Pareto optimal solution at any point on the trade-off curve. Specifically, to incorporate selection bias into our framework, we simply define

$$c(v) \triangleq |v - 1/2|, \quad \tau \triangleq n, \quad \mathcal{V}_k \triangleq [0,1]. \quad (16)$$

Our approach can consequently produce any design on the Pareto frontier, and thus Pareto dominate state-of-the-art BCD designs. We will see this numerically in Section 6.

Notice that the optimal policy Equation (15) in the setting of Equation (16) is a linear program. Direct examination of this program yields an interesting insight: at every time k, the optimal action for (P3′′) is restricted to $v_k \in \{0, 1/2, 1\}$. In other words, an optimal policy will only either take a deterministic action or fully randomize. This is in contrast to the main BCD heuristics developed in the literature (some of which we will describe shortly in Section 6.3), which tend to vary probabilities over the entire interval $[0,1]$.

5.2. Allocation Budget

Assuming a test with a total sample size of n, the designer may be happy to assign these samples to the control arm (the status quo) but may want to limit exposure to the test. Formally, we may want to have a budget B on the number of +1 allocations in the trial. As it turns out, BCD does not naturally extend to this setting (Han et al. 2009, Kuznetsova and Tymofyeyev 2012). Problem (P3′′) can trivially incorporate a budget constraint, we simply define

$$c(v) \triangleq 0, \quad \tau \triangleq n,$$
$$\mathcal{V}_k\left(\delta_{k-1}, \|\Delta_{k-1}\|^2_{\Sigma^{-1}}\right) \triangleq \begin{cases} [0,1] & \text{if } k + \delta_{k-1} < 2B, \\ \{0\} & \text{otherwise.} \end{cases}$$

5.3. Endogenous Stopping

Consider the (not uncommon) scenario where there is an economic cost associated with every incremental sampling unit in a sequential trial, and all we care about is estimating the treatment effect up to a desired level of precision (see Johari et al. 2017 for a broader discussion of related problems). In such a scenario, we may opportunistically want to stop early so that the sample size is in fact picked endogenously. For concreteness, let us suppose that the unit cost per sample is a constant r. Assume further that it suffices to estimate the treatment effect with precision κ, unless the trial has run up to a sample size of n, in which case we must stop. One can think of n here as an upper bound on sample size imposed by the trial designer. The objective is simply to minimize the expected cost of the trial. This problem is easily modeled in our framework. Specifically, (P3′′) can capture this problem by defining

$$c(v) \triangleq r, \quad \tau \triangleq \min\left\{k \geq 1 : k - \frac{1}{k}\left(\delta_k^2 + \|\Delta_k\|^2_{\Sigma^{-1}}\right) \geq \kappa\sigma^2\right\} \wedge n, \quad \mathcal{V}_k \triangleq [0,1].$$

6. Experiments

This section focuses on numerical experiments with data. We will attempt to highlight the relative merits of our approach vis-à-vis simple randomization, as well as BCDs. As discussed in the literature review, BCDs are an approach to minimizing loss (or equivalently, maximizing efficiency) by dynamically adjusting for covariate imbalances.

Our goal will be to show that for a given level of selection bias, our approach provides an improvement in efficiency (or a reduction in loss) over competing BCDs. Equivalently, our approach can achieve

a given level of efficiency with a smaller level of selection bias. We will study these relative merits for varying values of sample size n, and the number of covariates p. Finally, whereas our analysis in Section 4 required the covariates to follow an elliptical distribution, such a requirement may not hold in real applications. As such we conduct experiments using click log data from Yahoo! wherein the covariates are categorical. We show that our approach enjoys similar relative merits in this setting.

6.1. BCDs, Loss, and Selection Bias

Let $v_k \in [0,1]$ denote the probability that the kth allocation is set to $x_k = +1$ under a given allocation rule $\mathscr{A}$. Recall from Section 5 that a measure of selection bias under $\mathscr{A}$ is defined according to

$$\mathrm{Bias}_{\mathscr{A}} \triangleq \mathsf{E}\left[\frac{2}{n}\sum_{k=1}^{n}|v_k - 1/2|\right] \in [0,1].$$

(Here, we have normalized the bias to be contained in the unit interval.) This measure captures the extent of randomness (or, equivalently, how predictable any given allocation is) under $\mathscr{A}$ (Blackwell and Hodges 1957). Also, recall our definition of loss:

$$\mathrm{Loss}_{\mathscr{A}} \triangleq n - \mathsf{E}\left[x^\top P_{Z^\perp} x\right] = \mathsf{E}\left[x^\top Z\left(Z^\top Z\right)^{-1} Z^\top x\right] \geq 0.$$

The loss under $\mathscr{A}$ is interpreted as the effective number of samples on which information is lost due to an imbalance in covariates. It is well known that any allocation rule engenders a trade-off between loss and selection bias, so that a comparison between rules ideally compares the entire trade-off curve attained by the two rules (Atkinson 2002). We will do precisely this in the experiments that follow.

Observe that the expressions for bias and loss do not depend on the experimental outcomes $\{y_k\}$. From an empirical perspective, this is helpful: we can assess any rule $\mathscr{A}$, given only access to the covariate distribution. The conclusions we draw on the relative merits of one approach with respect to another hold across any linear model for the given covariate structure.

6.2. Data

We run our experiments on two different data distributions for the covariates. Assumption 3 holds in both cases. Thus, $\{Z_k\}$ are i.i.d. and $Z_{k,1}$ is assumed to be 1. We run our experiments with the following sampling distributions for $Z_{2:p}$.

6.2.1. Synthetic Gaussian Data. In our synthetic experiments, we assume that $Z_{2:p}$ follows multivariate normal distribution. This is, of course, an elliptical distribution, so that Assumption 2 is satisfied. For the covariance matrix Σ, we set $\Sigma_{ii} = 1.0$ and $\Sigma_{ij} = 0.1$ for any $j \neq i$.

6.2.2. Yahoo! User Data. To experiment on data from a more realistic setting, we use a data set of user click log data from the Yahoo! front page.[11] The users here are visitors to "Featured Tab of the Today Module" on the Yahoo! front page. In the data set, each user has 136 associated features, such as age and gender. Each feature is binary, taking values in $\{0,1\}$. Some of these features were constant throughout the data set, and these were discarded. Duplicate and colinear features were discarded as well. Features were selected at random until up to $p = 40$ features were collected. Feature selection was repeated independently in each simulation trial.

Our algorithm requires the covariance matrix of the data as an input. For this purpose, we estimate the covariance matrix from a portion of the data set. This estimate is obtained by simply taking a sample average across 1 million data points kept aside from the rest of the experiments.

Finally, for evaluation purposes, we require a generative model for the data. To this end, from a set of 1 million data points we sample individual data points, with replacement. In other words, as the sampling distribution we use the empirical distribution of the 1 million data points used for testing. Such a sampling procedure is intended to mimic the arrival of users on the Yahoo! front page.

6.3. Algorithms

6.3.1. Dynamic Programming (Our Approach). The problem at hand is addressed by the dynamic programming formulation described in Section 5. As such, we are required to compute the two-dimensional value functions given by $\{q_k\}_{1\leq k\leq n}$. These functions are computed offline by backward induction following Equation (14). Here, we provide the computational details for this operation. In particular, given $q_{k+1}(\cdot,\cdot)$, we compute $q_k(\cdot,\cdot)$ as follows:

1. Discretization: The first state variable m is discrete and can take values from $-n$ to n. We discretize values for the second state variable λ on a geometric mesh taking values λ_0^i for $\lambda_0 \triangleq 1.5$ and $0 \leq i \leq 26$. The maximum value of λ was chosen so that $\|\Delta_k\|^2_{\Sigma^{-1}}$ has a low probability of exceeding it.

2. Sampling: For each discretized pair (m,λ) we estimate $q_k(m,\lambda)$ via Monte Carlo simulation. In particular, $N = 10{,}000$ pairs[12] $(\xi,\eta) \in \mathbb{R}^2$ are sampled from the appropriate distributions and $q_k(m,\lambda)$ is estimated according to Equation (14) using the corresponding empirical measure. We use the same sample set of (ξ,η) for all (m,λ) at which this is evaluated.

3. Interpolation: Given an (m,λ) such that λ is not a discretized mesh point, we estimate $q_{k+1}(m,\lambda)$ in the Bellman recursion (Equation (14)) by linear interpolation between the closest points in the discretized mesh.

6.3.2. Biased Coin Designs. In addition to our own dynamic programming algorithm, we will consider several other rules proposed in the literature. These include Rule ABCD (Baldi Antognini and Zagoraiou 2011), which, following Atkinson (2014), we refer to as Rule J; Smith's rule (Rule S) (Smith 1984a, b); Atkinson's rule (Rule A) (Atkinson 1982); and the Bayesian procedure of Ball et al. (1993) (Rule B). Rules J, S, and B are all parameterized by a scalar parameter, which we denote ρ, that may take values in $(0, \infty)$. Rule A is a special case of Rule S taking $\rho = 1$. As $\rho \to 0$, these rules become equivalent to randomization. On the other hand, as $\rho \to \infty$, these rules become entirely deterministic in nature. As such, for values of ρ close to zero, one expects low selection bias, whereas as $\rho \to \infty$ one expects to see a reduction in loss at the expense of selection bias. A deterministic rule has the largest possible selection bias of 1. To precisely specify each of these rules, define

$$d_k\left(u_{k+1}, Z_{k+1,2:p}\right) \triangleq \left(1 - u_{k+1}\delta_k/k - u_{k+1}Z_{k+1,2:p}^{\top}\Sigma^{-1}\Delta_k/k\right)^2,$$

where $u_{k+1} \in \{\pm 1\}$, $Z_{k+1,2:p} \in \mathbb{R}^{p-1}$, and δ_k and Δ_k have the usual definitions (Equation (4)). For background on the function $d_k(\cdot,\cdot)$, see Atkinson (1982); this quantity arises naturally in the sequential design of D_A-optimal experiments. The rules described above then take the following form:

1. Rules S/A: Assign $x_{k+1} = +1$ with probability

$$v_{k+1} \triangleq \frac{d_k\left(+1, Z_{k+1,2:p}\right)^{\rho}}{d_k\left(+1, Z_{k+1,2:p}\right)^{\rho} + d_k\left(-1, Z_{k+1,2:p}\right)^{\rho}}.$$

The parameter ρ can take values in $(0, \infty)$. Rule A corresponds to the special case where $\rho = 1$.

2. Rule B: Assign $x_{k+1} = +1$ with probability

$$v_{k+1} \triangleq \frac{\left(1 + d_k\left(+1, Z_{k+1,2:p}\right)\right)^{\rho}}{\left(1 + d_k\left(+1, Z_{k+1,2:p}\right)\right)^{\rho} + \left(1 + d_k\left(-1, Z_{k+1,2:p}\right)\right)^{\rho}}.$$

The parameter ρ can again take values in $(0, \infty)$. This rule is very similar to Rule S, but permits a Bayesian interpretation (Ball et al. 1993).

3. Rule D: Assign $x_{k+1} = +1$ deterministically if $d_k(+1, Z_{k+1,2:p}) > d_k(-1, Z_{k+1,2:p})$, set $x_{k+1} = -1$ otherwise. This rule is obtained in the limit as $\rho \to \infty$ for rules A, S, and B. Note that this deterministic rule is equivalent to a myopic policy that seeks to optimize the objective of (P3') assuming that x_{k+1} is the final allocation to be made, and ignoring the impact of this allocation on future decision making.

4. Rule J: Define the discrepancy after k allocations, $D_k(Z_{k+1,2:p})$, according to

$$D_k(Z_{k+1,2:p}) \triangleq \frac{2 - k\left(d_k\left(+1, Z_{k+1,2:p}\right) + d_k\left(-1, Z_{k+1,2:p}\right)\right)}{d_k\left(+1, Z_{k+1,2:p}\right) - d_k\left(-1, Z_{k+1,2:p}\right)},$$

assuming $d_k(+1, Z_{k+1,2:p}) \neq d_k(-1, Z_{k+1,2:p})$. If $D_k \cdot (Z_{k+1,2:p}) < 0$, we assign $x_{k+1} = +1$ with probability

$$v_{k+1} \triangleq \frac{\left|D_k\left(Z_{k+1,2:p}\right)\right|^{\rho}}{1 + \left|D_k\left(Z_{k+1,2:p}\right)\right|^{\rho}}.$$

If, on the other hand $D_k(Z_{k+1,2:p}) > 0$, we assign $x_{k+1} = +1$ with probability

$$v_{k+1} \triangleq \frac{1}{1 + \left|D_k\left(Z_{k+1,2:p}\right)\right|^{\rho}}.$$

Finally, if $D_k(Z_{k+1,2:p}) = 0$ or $d_k(+1, Z_k) = d_k(-1, Z_k)$, we simply randomize ($v_{k+1} = 1/2$). The parameter ρ can again take values in $(0, \infty)$.

6.4. Results

Our goal is to compare the statistical efficiency of our dynamic programming-based sequential algorithm to the various competing BCDs discussed earlier while controlling for selection bias. To do this, we run each BCD procedure for an increasing sequence of value of ρ. The smallest value used, $\rho = 0$, is simply equivalent to randomized allocation. The largest value of ρ we considered for each scheme was chosen so that the rule was effectively deterministic. We implemented our sequential DP algorithm for an increasing sequence of values of γ, tracing out a similar trade-off curve.

Results are reported in Figures 1, 2, and 3. Of these, Figures 1 and 2 show results on synthetic Gaussian data whereas Figure 3 shows results on the Yahoo! data set. Each data point in these figures is the average of 10,000 independent Monte Carlo trials with shared randomness across all BCD rules and our own rule; and different data points were generated for each rule by varying their respective configurations of ρ and γ.

These figures reveal that:

1. For any target level of selection bias, our dynamic programming algorithm has the smallest loss among all of the alternatives implemented. In this way, the DP approach Pareto dominates all alternatives. The relative improvement in loss can be nontrivial: the loss incurred under our approach can be up to five times smaller for moderate budgets on

Figure 1. (Color online) Bias-Loss Trade-Off on Synthetic Gaussian Data for $n = 100$ and Varying Values of p

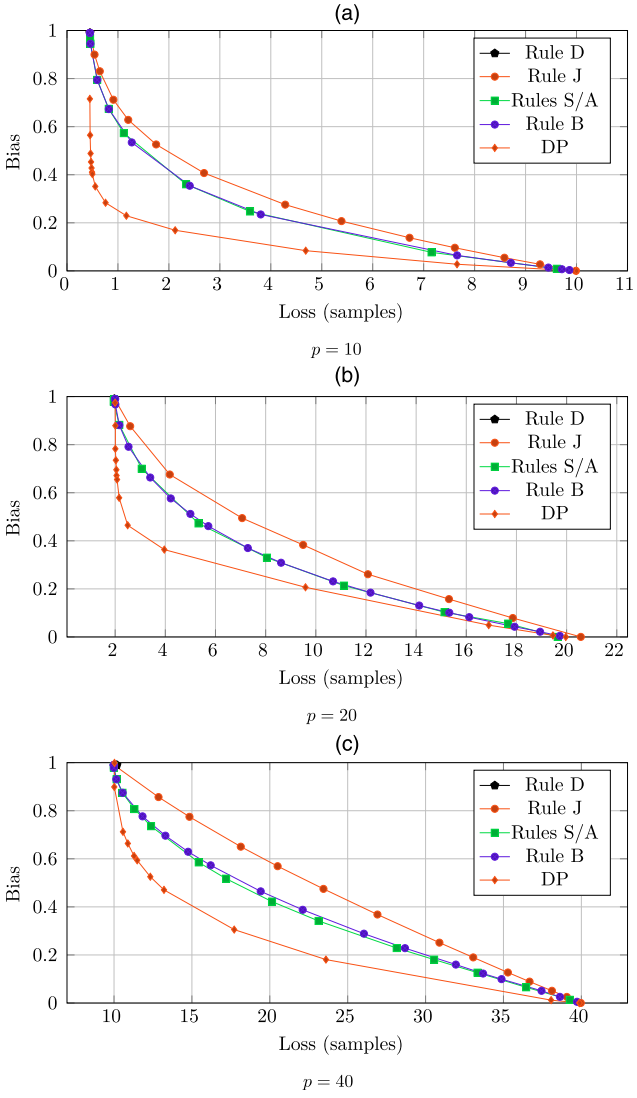
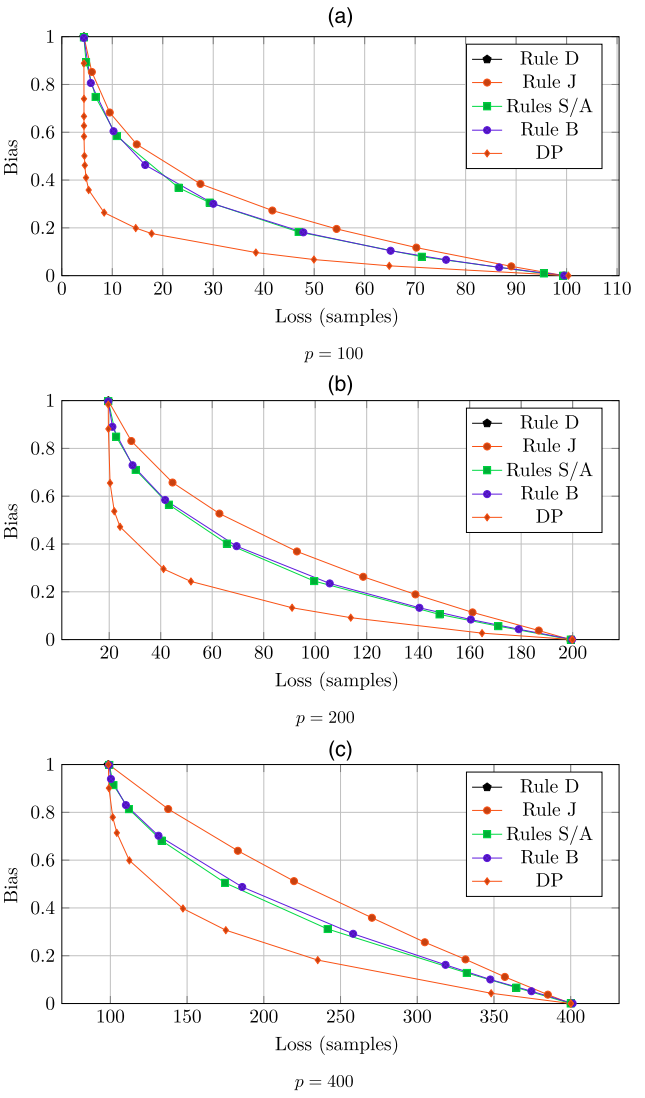


Figure 2. (Color online) Bias-Loss Trade-Off on Synthetic Gaussian Data for $n = 1{,}000$ and Varying Values of p



selection bias. Put a different way, the effective number of samples lost due to covariate imbalance can be substantially smaller for a given budget on selection bias .

2. The relative improvement alluded to previously is particularly pronounced for smaller values of p/n. Our intuition here is as follows: keeping n fixed, one expects to require fewer nonrandom allocations for small p. As such, the importance of strategizing on when to employ a nonrandom allocation has greater impact in such a setting.

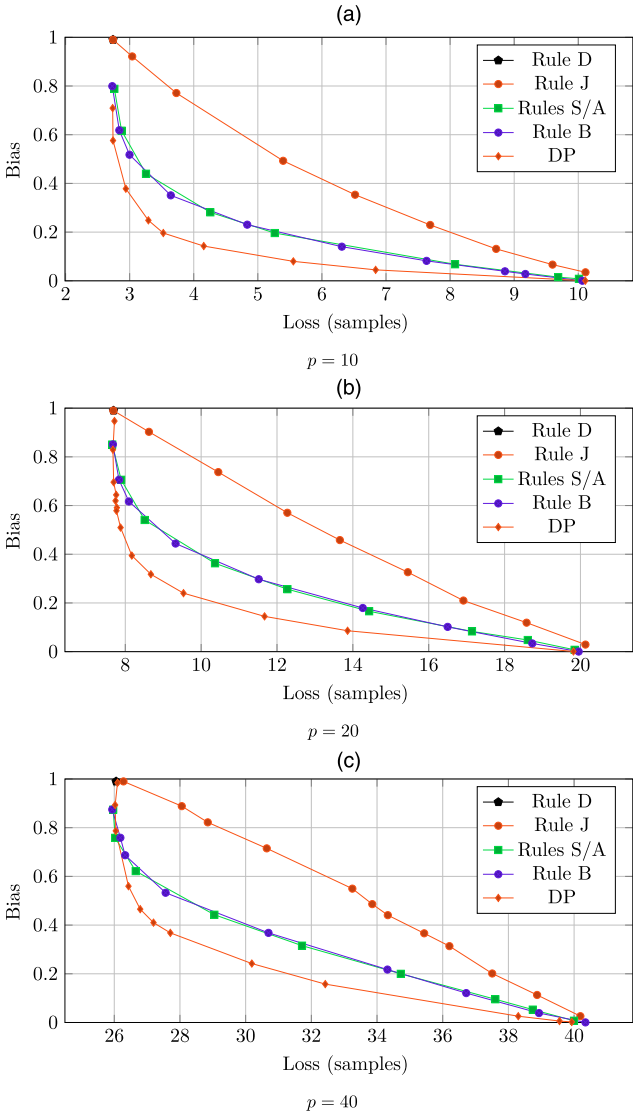
3. The relative merits of our sequential approach appear more pronounced in the setting where n is larger.

4. Finally, observe that Figure 3 shows results on the Yahoo! data set, and that the covariates in this experiment are in fact categorical. Despite this we see that our approach exhibits similar improvements relative to the competing BCD schemes.

7. Conclusion

We conclude with a summary of what we have accomplished and what we view as key directions for further research. At a conceptual level, this paper illustrates the power of the optimization viewpoint in what are inherently statistical problems: we have presented a provably near-optimal solution to a problem for which a plethora of heuristics were available. In addition to establishing the appropriate approach to this problem, the algorithms we have developed are eminently practical and easy to implement—a property that is crucial for the sorts of applications that motivated this work. On a more pragmatic note, we have quantified the value of these sorts of optimization approaches establishing precise estimates of the benefits optimization approaches provide over straightforward randomization. These estimates illustrate that in so-called high-dimensional setting—that is, in settings where the number of

Figure 3. (Color online) Bias-Loss Trade-Off on the Yahoo! Data Set for $n = 100$ and Varying Values of p



covariates is large, such approaches can provide order of magnitude improvements in sampling efficiency.

Our progress does come at the expense of structural assumptions on the relationship between the observed effect and observable covariates. In particular, we assumed a linear model with exogenous noise. Any such structural assumption is restrictive. In the event that these assumptions fail, they could result in biased estimates of the treatment effect. With that said, it appears difficult to overcome the risk of such a bias while using a covariate dependent treatment assignment scheme. In addition to these structural assumptions, we also required that the experiment designer have some knowledge on the distribution of the covariates (their covariance matrix). Our theoretical results made further distributional assumptions on these covariates. Much remains to be done to mitigate the impact of these limiting assumptions, and as such a number of directions remain for future research. We highlight several here in parting:

1. Normality: To what extent can our assumption on the normality of covariates be relaxed? Can we develop approximation guarantees for the situation when covariates are not normally distributed?

2. Nonlinear models: Can we allow for a nonlinear dependence on covariates? One direction to accomplish this is perhaps a reliance of some manner of nonparametric kernel approach. The good news here is that the value of optimization is likely to be even higher in such an infinite-dimensional setting.

3. More than two alternatives: The present paper considers only the two alternative setting, an important direction for future work would be to consider settings where there is a larger number of choices.

Acknowledgments

The authors thank Anthony Atkinson, Steve Chick, Shane Hendersen, Nathan Kallus, Costis Maglaras, Assaf Zeevi, and Jose Zubizarreta for helpful discussions.

Endnotes

[1] Rule C in table 1 of Atkinson (2002) illustrates that with as few as 10 covariates, methods based on stratification are hardly better than randomization.

[2] For example, consider the following case study by one of the largest providers of commercial A-B testing infrastructure: https://blog.optimizely.com/2014/02/03/case-study-sony-ab-tests-banner-ads/, accessed July 29, 2019.

[3] The assumption of no endogeneity is required for the least square estimate of θ under a given allocation to be unbiased. It is also required for our performance analysis. In general, it appears difficult to overcome bias in the face of the risk of model misspecification while using a covariate dependent treatment assignment scheme.

[4] We will informally refer to p as the number of covariates even though, strictly speaking, it is the dimension of the linear model and could include second order terms, interaction terms between covariates, and so on.

[5] Note that, in the online case, because of Jensen's inequality, maximizing precision and minimizing variance are no longer equivalent objectives.

[6] In what follows, given a function $f(\cdot)$ and a positive function $g(\cdot)$, as $n \to \infty$ we say $f(n) = O(g(n))$ if $\limsup_{n\to\infty} |f(n)|/g(n) < \infty$; we say $f(n) = o(g(n))$ if $\lim_{n\to\infty} |f(n)|/g(n) = 0$; we say $f(n) = \Omega(g(n))$ if $\limsup_{n\to\infty} |f(n)/g(n)| > 0$; and finally we say $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

[7] Here, $Y \succeq 0$ denotes that Y is a symmetric and positive semi-definite matrix.

[8] Note that under our assumption, it is easy to verify that each covariate vector $Z_{k,2:p}$ is zero mean. Our choice of normalization $\mathsf{E}[R^2] = p - 1$ ensures that the covariance matrix of $Z_{k,2:p}$ is given by Σ. This second moment requirement does exclude heavy-tailed elliptical distributions such as the Cauchy distribution. However, it is necessary so that our performance criteria (expected precision) is finite.

[9] If $p = 2$, we take $\xi \triangleq 0$.

[10] For the decomposition (Equation (14)) to apply, an additional technical assumption is needed on the stopping time τ: we assume that, for each $1 \le k < \tau$, the distribution of the random

variable corresponding to the future stopped payoff $\delta_\tau^2 + \|\Delta_\tau\|^2_{\Sigma^{-1}}$ is conditionally independent of the history given the current state (δ_k, Δ_k).

[11] This data set is obtained from the Yahoo! Labs repository of data sets available for academic research, and can be downloaded as "R6B—Yahoo! Front Page Today Module User Click Log Dataset, version 2.0" at http://webscope.sandbox.yahoo.com/catalog.php?datatype=r, accessed July 29, 2019.

[12] In all examples, our algorithm assumes that the covariate data are generated from a multivariate normal, even when this was not true (Yahoo! data set). In this case, when $Z_{2:p} \sim N(I, \Sigma)$ is multivariate normal, $\lambda + 2uRU_1\sqrt{\lambda} + R^2$ has the same distribution as $(\sqrt{\lambda} + u\eta)^2 + \xi$, where η is a standard normal and ξ is a chi-squared random variable with $p-2$ degrees of freedom. See also Corollary 2.

References

Atkinson AC (1982) Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* 69(1):61–67.

Atkinson AC (1999) Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statist. Medicine* 18(14):1741–1752.

Atkinson AC (2002) The comparison of designs for sequential clinical trials with covariate information. *J. Royal Statist. Soc. Ser. A* 165(2):349–373.

Atkinson AC (2014) Selecting a biased-coin design. *Statist. Sci.* 29(1): 144–163.

Baldi Antognini A, Zagoraiou M (2011) The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika* 98(3):519–535.

Ball FG, Smith AFM, Verdinelli I (1993) Biased coin designs with a Bayesian bias. *J. Statist. Planning Inference* 34(3):403–421.

Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* (SIAM, Philadelphia).

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

Blackwell D, Hodges JL (1957) Design for the control of selection bias. *Ann. Math. Statist.* 28(2):449–460.

Cambanis S, Huang S, Simons G (1981) On the theory of elliptically contoured distributions. *J. Multivariate Anal.* 11(3):368–385.

Chick SE, Frazier P (2012) Sequential sampling with economics of selection procedures. *Management Sci.* 58(3):550–569.

Chick SE, Gans N (2009) Economic analysis of simulation selection problems. *Management Sci.* 55(3):421–437.

Chick SE, Forster M, Pertile P (2017) A Bayesian decision theoretic model of sequential experimentation with delayed response. *J. Royal Statist. Soc. Ser. B Statist. Methodology* 79(5):1439–1462.

Cook TD, Campbell DT, Day A (1979) *Quasi-Experimentation: Design & Analysis Issues for Field Settings* (Houghton Mifflin, Boston).

Efron B (1971) Forcing a sequential experiment to be balanced. *Biometrika* 58(3):403–417.

Fisher RA (1935) *The Design of Experiments* (Oliver & Boyd, Edinburgh, UK).

Goemans MX, Williamson DP (1995) Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM* 42(6):1115–1145.

Han B, Enas NH, McEntegart D (2009) Randomization by minimization for unbalanced treatment allocation. *Statist. Medicine* 28(27):3329–3346.

Hauser JR, Urban GL, Liberali G, Braun M (2009) Website morphing. *Marketing Sci.* 28(2):202–223.

Hu Y, Hu F (2012) Asymptotic properties of covariate-adaptive randomization. *Ann. Statist.* 40(3):1794–1815.

Johari R, Pekelis L, Walsh DJ (2017) Always valid inference: Bringing sequential analysis to A/B testing. Working paper, Stanford University, Stanford, CA.

Kallus N (2013) Regression-robust designs of controlled experiments. Working paper, Cornell Tech, New York.

Kapelner A, Krieger A (2014) Matching on-the-fly: Sequential allocation with higher power and efficiency. *Biometrics* 70(2):378–388.

Kasy M (2016) Why experimenters should not randomize, and what they should do instead. Working paper, Harvard University, Cambridge, MA.

Kim S-H, Nelson BL (2006) Selecting the best system. Henderson S, Nelson B, eds. *Handbooks in Operations Research and Management Science: Simulation*, vol. 13 (North Holland, Amsterdam), 501–534.

Kuznetsova OM, Tymofyeyev Y (2012) Preserving the allocation ratio at every allocation with biased coin randomization and minimization in studies with unequal allocation. *Statist. Medicine* 31(8):701–723.

Langford J, Zhang T (2007) The epoch-greedy algorithm for contextual multi-armed bandits. *Advances in Neural Information Processing Systems* (Curran Associates, Red Hook, NY), 1096–1103.

Nesterov Y (1997) Semidefinite relaxation and nonconvex quadratic optimization. Technical report, Center for Operations Research and Econometrics, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Pocock SJ, Simon R (1975) Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 31(1):103–115.

Pukelsheim F (2006) *Optimal Design of Experiments* (SIAM, Philadelphia).

Raudenbush SW, Martinez A, Spybrook J (2007) Strategies for improving precision in group-randomized experiments. *Ed. Evaluation Policy Anal.* 29(1):5–29.

Rosenberger WF, Sverdlov O (2008) Handling covariates in the design of clinical trials. *Statist. Sci.* 23(3):404–419.

Schwartz EM, Bradlow ET, Fader PS (2017) Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Sci.* 36(4):471–643.

Smith RL (1984a) Properties of biased coin designs in sequential clinical trials. *Ann. Statist.* 12(3):1018–1034.

Smith RL (1984b) Sequential treatment allocation using biased coin designs. *J. Royal Statist. Soc. Ser. B* 46(3):519–543.

Steensma DP, Kantarjian HM (2014) Impact of cancer research bureaucracy on innovation, costs, and patient care. *J. Clinical Oncology* 32(5):376–378.

Toubia O, Hauser JR, Simester DI (2004) Polyhedral methods for adaptive choice-based conjoint analysis. *J. Marketing Res.* 41(1): 116–131.

Toubia O, Simester DI, Hauser JR, Dahan E (2003) Fast polyhedral adaptive conjoint estimation. *Marketing Sci.* 22(3):273–303.

Woodroofe M (1979) A one-armed bandit problem with a concomitant variable. *J. Amer. Statist. Assoc.* 74(368):799–806.