Effective algorithms for constructing optimal two-armed designs in the presence of subject's information

Alan R. Vazquez[1] and José Fernando Camacho-Vallejo[2]

[1]Department of Statistics, University of California, Los Angeles, U.S.A.
[2]Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Nuevo León, México

September 9, 2021

1 Introduction

Clinical trials are experiments that evaluate new treatments for diseases such as cancer. These trials are generally expensive due to the costs associated with recruiting the subjects, defining the locations of the testing facilities and implementing the treatment protocol. Therefore, it is important to plan clinical trials well.

Clinical trials can be categorized into three phases, each of which addresses a specific objective. For instance, Phase I trials find the safe dose range of a new treatment, while Phase II trials identify the treatment's dose level with the highest response between two or more potential options. Once the the optimal dose level of the new treatment is determined, Phase III trials compare its performance with the standard treatment using two-armed experimental designs. The goal is to demonstrate that the former outperforms the latter in terms of improving the subjects' quality of life, measured by one or more outcomes or responses. This last phase of clinical trials is the subject of this document.

A two-armed experimental design assigns one of the two treatments to each subject in the study. Ideally, the design collects high-quality data that allows to answer the research questions. In technical terms, we assume that the link between the treatments and the responses can be represented using a statistical model, and that the research question involves testing for the significance of one or more of this model's parameters. An optimal two-armed design then maximizes the estimation efficiency of these parameters.

When recruiting the subjects of a clinical trial, it is common to record information such as their age, gender, weight and height. Traditionally, the construction and analysis two-armed designs does not use this information. However, there is a recent trend to incorporate the interaction between these covariates and the treatments in the decision making for a subject. This trend is due to the surge of precision medicine, which seeks to maximize the quality of health-care by providing individual-level treatment for each subject. Kosorok and Laber (2019) show that incorporating subject individualized information when prescribing a treatment can boost its efficacy.

Atkinson (2015), Bertsimas et al. (2015) and Zhang et al. (2021) propose methods to incorporate subject-specific information in the construction of optimal two-armed designs. More specifically, Atkinson (2015) constructs optimal designs that maximize the estimation efficiency of a linear model containing the treatment effects and the interactions between treatments and covariates, in which the errors are heteroscedastic. Instead of showing the specific treatment to be assigned to each patient, these designs show the proportions of subjects that have to be assigned to each treatment. Bhat et al. (2020) construct optimal designs for a model which ignores the interactions between the treatment and covariates. The goal of their approach is to produce efficient estimates of the treatment effects. To this end, these authors develop an effective algorithm rooted in semidefinite programming. Bertsimas et al. (2015) introduces a mixed integer optimization model that finds the optimal allocation of subjects to treatments, so that the mean and variances of the groups' covariates are as similar as possible. Under this allocation, these authors propose a statistical test for the differences between the two treatment effects.

In contrast with Atkinson (2015), Bertsimas et al. (2015), and Bhat et al. (2020), Zhang et al. (2021) addressed the problem of allocating the treatments to each subject, so as to optimize the predicted response for subjects in and out of the sample. In this document, we describe the methods of these authors, which are rooted in bi-level optimization, and explore some interesting research ideas.

The rest of this document is organized as follows. Section 2 presents an overview of bi-level optimization and some important issues on the solution methods employeed. Section 3.1 introduces the statistical model for two-armed design as well as the goal of personalized medicine. Section 4 shows the optimal design problem which can be formulated as a bi-level programming problem.

2 Background on bi-level optimization

In real-life, there are several situations that involve two interrelated decision-making levels under a predefined hierarchy between them. From a game theory point-of-view, hierarchized two-level problems are named as Stackelberg's game Von Stackelberg (2010). The upper-level problem is associated with a leader, while the lower-level problem is associated with a follower. In this game, both decision makers has his/her own objective function to optimize. Leader's decisions affect the follower's decision space. On the other hand, the follower acts in a rational manner, and his/her decisions affect the leader's objective function and -sometimes- his/her decision space. First, these two-level hierarchized problems were studied as mathematical programming problems with an optimization problem within its constraints Bracken and McGill (1973). Later, the notation of bi-level programming was introduced in Candler and Norton (1977). Diverse interesting applications of bi-level problems had been summarized in Bard (2013), Colson et al. (2007), Labbé and Violin (2013), and Kalashnikov et al. (2015).

A bi-level program consists of a model in which the decision variables can be partitioned into two subsets, and one of these subsets will be determined by the optimal solution of another mathematical programming model. A general formulation of this problem is the one proposed by Dempe (2002) and it is shown next:

$$\min_{y \in Y} \quad F(x(y), y) \tag{1a}$$

$$s.t. \quad G(x(y), y) \leq 0, \tag{1b}$$

$$H(x(y), y) = 0, \tag{1c}$$

$$\text{where for a fixed } y, \text{ variable } x \text{ solves:} \tag{1d}$$

$$\min_{x \in X} f(x, y) \tag{1e}$$

$$s.t. \quad g(x, y) \leq 0, \tag{1f}$$

$$h(x, y) = 0, \tag{1g}$$

where $F : X \times Y \rightarrow \mathbb{R}$ and $f : X \times Y \rightarrow \mathbb{R}$. Furthermore, $G : X \times Y \rightarrow \mathbb{R}^k$ and $H : X \times Y \rightarrow \mathbb{R}^l$, $g : X \times Y \rightarrow \mathbb{R}^p$ and $h : X \times Y \rightarrow \mathbb{R}^q$, that is, $G(x, y) = (G_1(x, y), ..., G_k(x, y))$, $H(x, y) = (H_1(x, y), ..., H_l(x, y))$, $g(x, y) = (g_1(x, y), ..., g_p(x, y))$, and $h(x, y) = (h_1(x, y), ..., h_p(x, y))$.

Leader's decision variables are denoted by x, while the follower's ones are y. In this general bi-level model, the leader aims to minimize the objective function $F(x, y)$ subject to constraints given by Equations (1b) and (1c). On the other hand, the follower aims to minimize his/her own objective function $f(x, y)$, subject to constraints given by Equations (1f) and (1g).

The most common exact method used for solving bi-level problems is based on the reformulation of the bi-level problem into an equivalent single-level one. The latter is achieved by using the sufficient and necessary KKT conditions (when the follower's problem is convex) and by considering the pertinent strategies to handle the non-linear constraints. However, the resulting reformulations tend to be intractable for medium and large-size instances due to the characteristics of the constraints or the number of variables included for the linearization.

Furthermore, bi-level programming problems are NP-hard Bard (1991) and Jeroslow (1985), even in the simplest case in which both decision levels are linear programming problems. The latter fact implies that does not exist an exact algorithm capable to solve any instance of the problem in polynomial time. Nevertheless, in recent years, there has been an increasing trend for the use of heuristics and metaheuristics to solve bi-level problems (e.g. Nucamendi-Guillén et al. (2018), Camacho-Vallejo and Garcia-Reyes (2019), Said et al. (2021), and A Tawhid and Paluck (2021)). Two main strategies when designing algorithms: using a nested approach or approximating the follower's optimal response. In the former, the follower's problem is optimally solved for each leader's decision Talbi (2013); while in the latter, the follower's reaction is approximated by a response method (see Sinha et al. (2017). The nested approach is commonly implemented since feasible bi-level solutions are obtained. In the response method approach, an additional final step is needed to guarantee that solutions belong to the inducible region. Nevertheless, both methods have numerically demonstrated effectiveness and efficiency when solving bi-level problems. Therefore, the use of metaheuristics is a viable option to solve bi-level problems in terms of quality of the objective function value and required computational time.

A research direction on the bi-level programming field that has gained a lot of attention, recently, is the multi-objective bi-level problems, see Calvete et al. (2021), Camacho-Vallejo et al. (2021), Abo-Elnaga and Nasr (2022). In these problems, the characterization of the Pareto front is aimed. Important adaptations of the common solution methods are needed to guide the search in the multi-objective bi-level problems. Nevertheless, the existing theory of multi-objective programming can be merged with the bi-level programming one to address appropriately these complex problems.

3 Background

3.1 The statistical model

To construct a two-armed design, we start from a sample of n subjects each of which has data on m covariates. The covariates can be quantitative or categorical. In what follows, we assume that each quantitative covariate is normalized to take values in the interval $[-1, 1]$. We also restrict our attention to categorical covariates which take two values only, coded as -1 and $+1$. We denote the space of feasible covariate vectors by $\mathcal{Z} = [-1, 1]^{m_q} \times \{-1, 1\}^{m_c}$, where m_q and m_c are the number of quantitative and categorical covariates, respecively, and $m = m_q + m_c$.

We denote the j-th covariate value of the i-th subject as z_{ij}, where $j = 1, \ldots, p$ and $i = 1, \ldots, n$. For this subject, we collect these values in the vector $\mathbf{z}_i = (z_{i1}, \ldots, z_{im})^T$.

To define the statistical model, we introduce the expansion function $\mathbf{h}(\mathbf{z})$, which takes a vector of covariates $\mathbf{z}$ and expands it to terms of interest. More specifically, $\mathbf{h}(\mathbf{z})$ is an $p \times 1$ vector that may contain the linear and interaction terms of the elements of $\mathbf{z}$, in addition to the quadratic terms of the quantitative covariates in $\mathbf{z}$. For example, consider a vector of quantitative covariates $\mathbf{z} = (z_1, z_2)^T$. The expansion function $\mathbf{h}(\mathbf{z}) = (z_1, z_2, z_1 z_2, z_1^2, z_2^2)^T$ contains the linear, quadratic and two-way interactions of z_1 and z_2. Without lost of generality, we assume that the first component of $\mathbf{h}(\mathbf{z})$ is equal to one.

We code the two treatments in the two-armed design as -1 and $+1$, and denote the treatment assigned to the i-th subject as x_i. The two-armed design is then given by the vector $\mathbf{x} = (x_1, \ldots, x_n)^T$. We denote the response of the i-th subject as y_i, which is assumed to take any real value.

The following model links the responses to the treatments and covariates of the subjects:

$$y_i = \mathbf{h}(\mathbf{z}_i)^T\boldsymbol{\alpha} + x_i\mathbf{h}(\mathbf{z}_i)^T\boldsymbol{\beta} + \epsilon_i, \tag{2}$$

where α_0 is the intercept, α_j is the effect of the j-th term of $\mathbf{h}(\mathbf{z}_i)$, β_0 is the treatment effect, β_j is the interaction effect between the treatment and the j-th term of $\mathbf{h}(\mathbf{z}_i)$, ϵ_i is the random error, $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p-1})^T$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})^T$. The random variables $\epsilon_1, \ldots, \epsilon_n$ are assumed to be independent and follow a normal distribution with zero mean and variance σ^2.

Model (2) includes the interaction between the treatment and the covariate terms. This is different from the model of Bhat et al. (2020) which does not contain these terms. Zhang et al. (2021) considered model (2) with $\mathbf{h}(\mathbf{z}_i)$ containing the linear terms of the covariates only. In contrast with these authors, we consider $\mathbf{h}(\mathbf{z}_i)$ to also include two-way interactions between the covariates, and the quadratic effects of the quantitative ones.

3.2 The goal of precision medicine

After collecting the responses from design $\mathbf{x}$, we estimate the vectors of parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ using the ordinary least squares estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$, respectively. The estimated model therefore is

$$\hat{y} = \mathbf{h}(\mathbf{z})^T\hat{\boldsymbol{\alpha}} + x\mathbf{h}(\mathbf{z})^T\hat{\boldsymbol{\beta}},$$

which can be used to predict the response of a subject with covariate vector $\mathbf{z} \in \mathcal{Z}$ (even if this subject is not part of the sample) under any of the two treatments.

The goal of personalized medicine is to recommend the treatment $x^\star$ to a subject so as to maximize the subject's response. More specifically, we need to find

$$x^\star := \operatorname*{arg\,max}_{x\in\{-1,+1\}} \mathbf{h}(\mathbf{z})^T\hat{\boldsymbol{\alpha}} + x\mathbf{h}(\mathbf{z})^T\hat{\boldsymbol{\beta}}. \tag{3}$$

for the subject with covariate vector $\mathbf{z}$. The relevant term in this problem is $x\mathbf{h}(\mathbf{z})^T\hat{\boldsymbol{\beta}}$, since it depends on the treatment assignment only.

Ideally, the data should allow us to estimate $x\mathbf{h}(\mathbf{z})^T\hat{\boldsymbol{\beta}}$ with maximum precision for all $\mathbf{z} \in \mathcal{Z}$. In other words, this estimate should have the smallest variance. Let $\mathbf{H} = (\mathbf{h}(\mathbf{z}_1), \ldots, \mathbf{h}(\mathbf{z}_n))^T$, $\mathbf{x} = (x_1, \ldots, x_n)^T$, and $\mathbf{D}_\mathbf{x} = \text{diag}(x_1, \ldots, x_n)$. Following the same arguments in Zhang et al. (2021), we can show that the variance of $x\mathbf{h}(\mathbf{z})^T\hat{\boldsymbol{\beta}}$ is $\sigma^2\mathbf{h}(\mathbf{z})^T\Sigma(\mathbf{x},\mathbf{H})\mathbf{h}(\mathbf{z})$, where

$$\Sigma(\mathbf{x},\mathbf{H}) = (\mathbf{H}^T\mathbf{H} - \mathbf{H}^T\mathbf{D}_\mathbf{x}\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{D}_\mathbf{x}\mathbf{H})^{-1} \tag{4}$$

is the $p \times p$ variance-covariance matrix of $\hat{\boldsymbol{\beta}}$. In our notation, we emphasize that $\Sigma(\mathbf{x},\mathbf{H})$ depends on the treatment vector $\mathbf{x}$ and the covariate values of the subjects in the sample, through $\mathbf{H}$.

3.3 Statistical criteria

Since matrix $\mathbf{H}$ is fixed, our goal is to minimize the variances $\sigma^2\mathbf{h}(\mathbf{z})^T\Sigma(\mathbf{x},\mathbf{H})\mathbf{h}(\mathbf{z})$ for all $\mathbf{z} \in \mathcal{Z}$, by a careful selection of the entries in the two-armed design $\mathbf{x}$. Inspired by the literature on optimal experimental design (Atkinson et al., 2007), we present the two criteria to summarize the variances for all the subjects in the population, given by a two-armed design. Our bi-level optimization approach minimizes these two criteria simultaneously.

3.3.1 G-optimality

The G-optimality criterion (Atkinson et al., 2007, ch. 10) is the maximum variance for all the covariate vectors in the feasible space. In technical terms, the G-optimality criterion of a two-armed design $\mathbf{x}$ is

$$\max_{\mathbf{z}\in\mathcal{Z}} \sigma^2\mathbf{h}(\mathbf{z})^T\ \Sigma(\mathbf{x},\mathbf{H})\ \mathbf{h}(\mathbf{z}). \tag{5}$$

The design $\mathbf{x}^\star$ that minimizes Equation (5) is referred to as a G-optimal design. The algorithms of Zhang et al. (2021) focus on minimizing the G-optimality criterion.

3.3.2 I-optimality

The I-optimality criterion (Atkinson et al., 2007, ch. 10) is the average variance for all the covariate vectors in the feasible space. For a two-armed design $\mathbf{z}$, the average variance is calculated as:

$$\frac{\int_\mathcal{Z}\ \sigma^2\mathbf{h}(\mathbf{z})^T\Sigma(\mathbf{x},\mathbf{H})\ \mathbf{h}(\mathbf{z})d\mathbf{z}}{\int_\mathcal{Z} d\mathbf{z}}. \tag{6}$$

The design $\mathbf{x}^\star$ that minimizes Equation (6) is referred to as an I-optimal design.

Calculating the I-optimality criterion is relatively easy. In Appendix A, we show that minimizing Equation (6), is equivalent to minimizing

$$\sigma^2 2^{-m} \operatorname{Tr}\left[\Sigma(\mathbf{x}, \mathbf{H})\, \mathbf{M}\right], \tag{7}$$

where $\mathbf{M}$ is the moments matrix. When the expansion function $\mathbf{h}(\mathbf{z})$ in model (2) contains linear, two-way interactions or quadratic terms, $\mathbf{M}$ has a very specific structure. Appendix A shows moments matrices for the expansion functions considered in this article.

4 Optimization problem

4.1 Problem formulation

Our optimal design problem is to find the vector $\mathbf{x}^\star$ that minimizes both the G- and I-optimality criteria. From an optimization perspective, we aim to solve the following bi-objective bi-level problem:

$$\min_{\mathbf{x},\, \mathbf{z}} \quad \left\{\mathbf{h}(\mathbf{z})^T\, \Sigma(\mathbf{x}, \mathbf{H})\, \mathbf{h}(\mathbf{z}),\ \ \operatorname{Tr}\left[\Sigma(\mathbf{x}, \mathbf{H})\, \mathbf{M}\right]\right\} \tag{8a}$$

subject to

$$x_i \in \{-1, 1\}, \quad i = 1, \ldots, n, \tag{8b}$$

$$\mathbf{z} := \arg\max_{\mathbf{z}} \mathbf{h}(\mathbf{z})^T\, \Sigma(\mathbf{x}, \mathbf{H})\, \mathbf{h}(\mathbf{z}) \tag{8c}$$

$$-1 \leq z_j \leq 1, \quad j = 1, \ldots, m_q, \tag{8d}$$

$$z_j \in \{-1, 1\}, \quad j = m_q + 1, \ldots, m. \tag{8e}$$

In this problem formulation, x_i is the i-th element of the vector $\mathbf{x}$, z_j is the j-th element of vector $\mathbf{z}$, m_q is the number of quantitative covariates in $\mathbf{z}$, and $\Sigma(\mathbf{x}, \mathbf{H})$ is given in Equation (4) with $\mathbf{H}$ fixed. This problem formulation involves $n + p$ decision variables contained within $\mathbf{x}$ and $\mathbf{z}$. The objective function in Equation (8a) is a vector which contains the G- and I-optimality criteria, ignoring the constant σ^2 because it is irrelevant in the optimization.

The problem formulation involves two types of constraints. The first type in Equation (8b) ensures that the variables x_i can take two possible values, -1 and 1. The second type of constraints is given by Equations (8c) - (8e) is in itself an optimization problem.

The solution to this problem is the vector $\mathbf{z}$ that maximizes the variance of $x\mathbf{h}(\mathbf{z})^T\hat{\boldsymbol{\beta}}$, ignoring the value of σ^2. The constraints of this inner problem in Equations (8d) define the upper and lower bounds of the quantitative covariates in $\mathbf{z}$. The contraints in Equations (8e) ensure that the categorical covariates in $\mathbf{z}$ can take two possible values, -1 and 1. Therefore, the constraints in Equations (8d) and (8e) define the space of feasible covariate vectors $\mathcal{Z}$.

The problem formulation in Equations (8a)–(8d) belongs to the class of multi-objective bi-level optimization problems, which are known to be NP-hard.

4.2 Step-by-step solution

A general procedure to solve the multi-objective bi-level optimization problem defined in the previous section is outlined in Algorithm 1. The sub-routine "Evaluate" computes the G- and I-optimality criterion values of a solution $\mathbf{x}$. The pseudocode of this subroutine is shown in Algorithm 2. The sub-routine "Change" makes a modification to an existing solution so as to construct a better one. Finally, the sub-routine "Update" updates a set of elements.

5 Numerical experiments

Using the approaches of Zhang et al. (2021), we generate minimax optimal designs for a sample of 100 and 250 subjects (n) with information on 6, 15 and 10 covariates (p). We generate five $n \times p - 1$ matrices $\mathbf{H}_u$ with elements -1 and $+1$ at random, $u = 1, \ldots, 5$. To assess an experimental design $\mathbf{x}$, we compute the maximum variance of $\sigma^2\mathbf{h}^T\Sigma(\mathbf{x}, \mathbf{H})\mathbf{h}$ over the whole space of binary covariate vectors. The computations were carried out using R code of these authors.

In addition to the exact and heuristic algorithms, Zhang et al. (2021) also report results on generating a design $\mathbf{x}$ at random. More specifically, they generated 100 designs $\mathbf{x}$ by randomly assigning $n/2$ -1s and $n/2$ $+1$s to their elements. Then, they choose the best one in terms of the criterion.

Table 1 shows the maximum variances obtained by the exact, heuristic and random algorithms, for the five randomly generated covariate matrices.

Algorithm 1: Pseudocode of algorithm for solving our multi-objective bi-level optimization problem.

Input: Maximum number of iterations M without improvement, moments matrix $\mathbf{M}$, and matrix of expansion functions of the covariate vectors of subjects in the sample $\mathbf{H}$.

Set $\mathcal{X} \leftarrow \varnothing$

Set $\mathcal{G} \leftarrow \varnothing$

Generate initial feasible solution $\mathbf{x}$ at random.

$\mathcal{X} \leftarrow \mathcal{X} \cup \mathbf{x}$

$\mathbf{G} \leftarrow \text{Evaluate}(\mathbf{x}, \mathbf{M}, \mathbf{H})$

$\mathcal{G} \leftarrow \mathcal{G} \cup \mathbf{G}$

Set $i \leftarrow 0$

while $i < M$ **do**

    $i \leftarrow i + 1$

    $\mathbf{x}' \leftarrow \text{Change}(\mathbf{x})$

    $\mathbf{G}' \leftarrow \text{Evaluate}(\mathbf{x}', \mathbf{M}, \mathbf{H})$

    if $\mathbf{G}'$ *dominates an element in* $\mathcal{G}$ **then**

        $\mathcal{G} \leftarrow \mathcal{G} \cup \mathbf{G}'$

        $\mathcal{X} \leftarrow \mathcal{X} \cup \mathbf{x}$

        $\mathbf{x} \leftarrow \mathbf{x}'$

        Update($\mathcal{G}$) #Remove dominated vectors

        Update($\mathcal{X}$) #Remove dominated solutions

        $i \leftarrow 0$

Output: Set of Pareto optimal solutions $\mathcal{X}$.

Appendix A: Selected moments matrices

Let the $m \times 1$ vector of covariates be $\mathbf{z} = (z_1, z_2, \ldots, z_{m_q}, z_{m_q+1}, \ldots, z_m)^T$ with m_q quantitative covariates and $m_c = m - m_q$ categorical covariates. In what follows, let $\mathbf{I}_n$ be the $n \times n$ identity matrix, $\mathbf{J}_{n\times m}$ be the $n \times m$ matrix of ones, $\mathbf{0}_{n\times m}$ be the $n \times m$ matrix of zeros, $\mathbf{0}_n$ be the $n \times 1$ vector of zeros, and $\mathbf{1}_n$ be the $n \times 1$ vector of ones.

We show the moments matrices for different expansion functions $\mathbf{h}(\mathbf{z})$ below.

Algorithm 2: Evaluate.

Input: Solution $\mathbf{x}$, moments matrix $\mathbf{M}$ and matrix $\mathbf{H}$.

1. $G_1 \leftarrow \mathrm{Tr}\left[\Sigma(\mathbf{x},\mathbf{H})\ \mathbf{M}\right]$
2. Set $\mathbf{z}_o$ to be the optimal solution of the problem in Equations (8c)–(8e).
3. $G_2 \leftarrow \mathbf{h}(\mathbf{z}_o)^T\ \Sigma(\mathbf{x},\mathbf{H})\ \mathbf{h}(\mathbf{z}_o)$
4. $\mathbf{G} \leftarrow (G_1, G_2)$

Output: Vector $\mathbf{G}$.

Case I. The expansion function contains the linear effects of all covariates.

In this case, we have that $\mathbf{h}(\mathbf{z}) = (1, \mathbf{z})$. It is easy to show that the moments matrix is:

$$\mathbf{M} = 2^m \begin{pmatrix} 1 & \mathbf{0}_{m_q}^T & \mathbf{0}_{m_c}^T \\ \mathbf{0}_{m_q} & \frac{1}{3}\mathbf{I}_{m_q} & \mathbf{0}_{m_q \times m_c} \\ \mathbf{0}_{m_c} & \mathbf{0}_{m_c \times m_q} & \mathbf{I}_{m_c} \end{pmatrix}. \tag{A1}$$

Case II. The expansion function contains the linear, quadratic or interaction effects of quantitative covariates.

Let $m = m_q$ and so, $\mathbf{z}$ has quantitative covariates only. In this case, we assume that the first element of $\mathbf{h}(\mathbf{z})$ is one. The following elements of the expansion function are the m linear effects, the m quadratic effects and the $m(m-1)/2$ interaction terms of the covariates. In this case, Goos and Jones (2011) shows that the moments matrix is:

$$\mathbf{M} = 2^m \begin{pmatrix} 1 & \mathbf{0}_m^T & \mathbf{0}_{m(m-1)/2}^T & \frac{1}{3}\mathbf{1}_m^T \\ \mathbf{0}_m & \frac{1}{3}\mathbf{I}_m & \mathbf{0}_{m\times m(m-1)/2} & \mathbf{0}_{m\times m} \\ \mathbf{0}_{m(m-1)/2} & \mathbf{0}_{m(m-1)/2\times m} & \frac{1}{9}\mathbf{I}_{m(m-1)/2} & \mathbf{0}_{m(m-1)/2\times m} \\ \frac{1}{3}\mathbf{1}_m & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m(m-1)/2} & \frac{1}{5}\mathbf{I}_m + \frac{1}{9}\left(\mathbf{J}_m - \mathbf{I}_m\right) \end{pmatrix}. \tag{A2}$$

References

A Tawhid, M. and Paluck, G. (2021). Solving linear bilevel programming via particle swarm algorithm with heuristic pattern searc. *Information Sciences Letters*, 6(1):1.

Abo-Elnaga, Y. and Nasr, S. (2022). K-means cluster interactive algorithm-based evolutionary approach for solving bilevel multi-objective programming problems. *Alexandria Engineering Journal*, 61(1):811–827.

Atkinson, A. (2015). Optimum designs for two treatments with unequal variances in the presence of covariates. *Biometrika*, 102:494–499.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press.

Bard, J. F. (1991). Some properties of the bilevel programming problem. *Journal of optimization theory and applications*, 68(2):371–378.

Bard, J. F. (2013). *Practical bilevel optimization: algorithms and applications*, volume 30. Springer Science & Business Media.

Bertsimas, D., Johnson, M., and Kallus, N. (2015). The power of optimization over randomization in designing experiments involving small samples. *Operations Research*, 63:751–978.

Bhat, N., Farias, V. K., Moallemi, C. C., and Sinha, D. (2020). Near-optimal A-B testing. *Management Science*, 66:4477–4495.

Bracken, J. and McGill, J. T. (1973). Mathematical programs with optimization problems in the constraints. *Operations Research*, 21(1):37–44.

Calvete, H. I., Galé, C., and Iranzo, J. A. (2021). An evolutionary algorithm for a bilevel biobjective location-routing-allocation problem. In *Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences*, pages 17–33. Springer.

Camacho-Vallejo, J.-F. and Garcia-Reyes, C. (2019). Co-evolutionary algorithms to solve hierarchized steiner tree problems in telecommunication networks. *Applied Soft Computing*, 84:105718.

Camacho-Vallejo, J.-F., López-Vera, L., Smith, A. E., and González-Velarde, J.-L. (2021). A tabu search algorithm to solve a green logistics bi-objective bi-level problem. *Annals of Operations Research*, pages 1–27.

Candler, W. and Norton, R. (1977). *Multi-level programming and development policy.* The World Bank.

Colson, B., Marcotte, P., and Savard, G. (2007). An overview of bilevel optimization. *Annals of operations research*, 153(1):235–256.

Dempe, S. (2002). *Foundations of bilevel programming*. Springer Science & Business Media.

Goos, P. and Jones, B. (2011). *Optimal Design of Experiments: A Case Study Approach*. Wiley.

Jeroslow, R. G. (1985). The polynomial hierarchy and a simple model for competitive analysis. *Mathematical programming*, 32(2):146–164.

Kalashnikov, V. V., Dempe, S., Pérez-Valdés, G. A., Kalashnykova, N. I., and Camacho-Vallejo, J.-F. (2015). Bilevel programming and applications. *Mathematical Problems in Engineering*, 2015.

Kosorok, M. R. and Laber, E. B. (2019). Precision medicine. *Annual Review of Statistics and Its Application*, 6(1):263–286.

Labbé, M. and Violin, A. (2013). Bilevel programming and price setting problems. *4OR*, 11(1):1–30.

Nucamendi-Guillén, S., Dávila, D., Camacho-Vallejo, J.-F., and González-Ramírez, R. G. (2018). A discrete bilevel brain storm algorithm for solving a sales territory design problem: a case study. *Memetic Computing*, 10(4):441–458.

Said, R., Elarbi, M., Bechikh, S., and Said, L. B. (2021). Solving combinatorial bi-level optimization problems using multiple populations and migration schemes. *Operational Research*, pages 1–39.

Sinha, A., Malo, P., and Deb, K. (2017). Evolutionary algorithm for bilevel optimization using approximations of the lower level optimal solution mapping. *European Journal of Operational Research*, 257(2):395–411.

Talbi, E.-G. (2013). A taxonomy of metaheuristics for bi-level optimization. In *Metaheuristics for bi-level optimization*, pages 1–39. Springer.

Von Stackelberg, H. (2010). *Market structure and equilibrium*. Springer Science & Business Media.

Zhang, Q., Khademi, A., and Song, Y. (2021). Mini-max optimal design of two-armed trials with side information. *INFORMS Journal of Computing*, Published online.

Table 1: Maximum variances over binary covariate vectors obtained by the exact, heuristic and random approach of Zhang et al. (2021).

n	p	Method	$\mathbf{H}_1$	$\mathbf{H}_2$	$\mathbf{H}_3$	$\mathbf{H}_4$	$\mathbf{H}_5$
100	6	HEURISTIC	0.0784	0.0879	0.0754	0.1016	0.0798
		EXACT	0.0925	0.0878	0.0752	0.1013	0.0809
		RANDOM	0.0788	0.0884	0.0757	0.1018	0.0806
	10	HEURISTIC	0.1667	0.1556	0.1554	0.1687	0.1377
		EXACT	0.1633	0.1554	0.1548	0.1650	0.1352
		RANDOM	0.1683	0.1574	0.1608	0.1725	0.1437
	20	HEURISTIC	0.5961	0.4682	0.4775	0.4632	0.4955
		EXACT	0.5796	0.4433	0.4751	0.4479	0.4817
		RANDOM	0.6394	0.5329	0.5551	0.5279	0.5818
150	6	HEURISTIC	0.0572	0.0532	0.0559	0.0490	0.0551
		EXACT	0.0563	0.0532	0.0559	0.0490	0.0551
		RANDOM	0.0564	0.0535	0.0562	0.0491	0.0551
	10	HEURISTIC	0.1066	0.0965	0.1130	0.0949	0.1117
		EXACT	0.1063	0.0961	0.1126	0.0945	0.1111
		RANDOM	0.1077	0.0974	0.1139	0.0972	0.1132
	20	HEURISTIC	0.2479	0.2440	0.2419	0.2546	0.2796
		EXACT	0.2421	0.2359	0.2436	0.2543	0.2755
		RANDOM	0.2640	0.2564	0.2586	0.2787	0.3025