Statistica Sinica **35** (2025), 1479-1498
doi:https://doi.org/10.5705/ss.202022.0414

CONSTRAINED D-OPTIMAL DESIGN FOR PAID RESEARCH STUDY

Yifei Huang[1], Liping Tong[2] and Jie Yang*[1]

[1] *University of Illinois at Chicago and* [2] *Advocate Aurora Health*

Abstract: We consider constrained sampling problems in paid research studies or clinical trials. When qualified volunteers are more than the budget allowed, we recommend a D-optimal sampling strategy based on the optimal design theory and develop a constrained lift-one algorithm to find the optimal allocation. Unlike the literature which mainly deals with linear models, our solution solves the constrained sampling problem under fairly general statistical models, including generalized linear models and multinomial logistic models, and with more general constraints. We justify theoretically the optimality of our sampling strategy and show by simulation studies and real-world examples the advantages over simple random sampling and proportionally stratified sampling strategies.

Key words and phrases: Constrained sampling, D-optimal design, generalized linear model, lift-one algorithm, multinomial logistic model.

1. Introduction

We consider a constrained sampling problem frequently arising in paid research studies or clinical trials, especially when recruiting volunteers via the Internet or emails, which could gather attention widely and quickly. For example, some investigators plan to conduct a research study to evaluate the effect of a new treatment on anxiety. Besides the treatment cost, the investigators also need to prepare certain compensation for participants' time. Due to limited funding, the investigators could only support up to n participants while there are $N > n$ eligible volunteers. The question is how they select n participants out of N to evaluate the treatment effect most accurately. Noted that the goal of the sampling problem in this paper is not the mean of response but the treatment effect or regression coefficients of an underlying statistical model.

A straightforward approach is to use the *simple random sampling without replacement* (SRSWOR) (Lohr, 2019, Ch. 2), which randomly chooses an index set $1 \le i_1 < i_2 < \cdots < i_n \le N$ such that each index set of n distinct subjects has the equal chance $n!(N-n)!/N!$ to be chosen. This can be applied if the investigators know nothing about the volunteers except contact information, or the covariate information provided by the volunteers does not seem relevant to the treatment effect.

*Corresponding author. E-mail: jyang06@uic.edu

A more common practice is, however, that the investigators collect some covariates information, such as gender and age, which may play some roles in the treatment effect. Suppose there are d covariates and m distinct combinations of covariates under consideration. For example, $d = 2$ covariates, gender (male or female), and age (18–25, 26–64, 65 and above), lead to $m = 6$ possible categories (combinations of covariates) of eligible volunteers, known as *strata* in the sampling theory (Lohr, 2019, Ch. 3). Suppose the frequencies of volunteers in the m categories are $N_1, \ldots, N_m$, respectively. The question is how we determine $n_i \leq N_i$ such that $n = \sum_{i=1}^{m} n_i$, known as the *allocation* of subjects to the categories or strata. Once an allocation $(n_1, \ldots, n_m)$ is determined, n_i subjects will be chosen randomly from the ith category or stratum for each i, known as a *stratified* (random) sampling. A commonly used stratified sampler chooses $n_i \propto N_i$, known as the *proportionally stratified* sampler, which is expected to produce a more accurate estimate for the mean response than SRSWOR (Lohr, 2019, Sec. 3.4.1).

Nevertheless, from an optimal design point of view (Fedorov, 1972; Silvey, 1980; Pukelsheim, 1993; Atkinson, Donev and Tobias, 2007; Fedorov and Leonov, 2014), we want to find n_i's such that the treatment effects or regression coefficients can be estimated most accurately. When no prior knowledge about the regression model is available, a *uniform* allocation, which assigns roughly the same number of subjects to each category, is commonly used in the practice of experimental design (Yang, Mandal and Majumdar, 2012). For the sampling problem under constraints $n_i \leq N_i$, $i = 1, \ldots, m$, we recommend (constrained) *uniformly stratified* sampler, which chooses $n_i = \min\{k, N_i\}$ or $\min\{k, N_i\} + 1$ with k satisfying $\sum_{i=1}^{m} \min\{k, N_i\} \leq n < \sum_{i=1}^{m} \min\{k+1, N_i\}$ (Sec. 4).

If the investigators have some information about the regression coefficients from some pilot study or prior research, we propose optimal stratified samplers based on the optimal design theory, which minimizes the variances of the estimated regression coefficients instead of the estimated population mean. According to different optimality criteria used (Fedorov, 1972; Atkinson, Donev and Tobias, 2007; Stufken and Yang, 2012; Fedorov and Leonov, 2014), we call the corresponding sampler D-optimal sampler, A-optimal sampler, etc. In this paper, we focus on D-optimality, which is the most frequently used (Zocchi and Atkinson, 1999; Atkinson, Donev and Tobias, 2007; Yang, Tong and Mandal, 2017).

In the statistical literature, optimal designs under constraints were considered mainly for linear models with the information $\mathbf{F}_i = \mathbf{h}(\mathbf{x}_i)\mathbf{h}(\mathbf{x}_i)^T$ obtained at the ith experimental setting $\mathbf{x}_i$ (Sec. 2), where $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), \ldots, h_p(\mathbf{x}))^T$ are known predictor functions; see Elfving (1952), Lee (1988), Cook and Fedorov (1995), Fedorov and Leonov (2014), and references therein. Among them, Wynn (1977a,b, 1982) connected finite population sampling with optimal designs under constraints $n_i \leq N_i$ (Example 1); and Welch (1982), Fedorov (1989), and Pronzato (2004, 2006) developed algorithms searching for "optimum submeasures" or

"optimum bounded designs". In this paper, $\mathbf{F}_i$ could be much more complicated and depend on unknown parameters $\boldsymbol{\theta}$, such as $\nu\{\mathbf{h}(\mathbf{x}_i)^T\boldsymbol{\theta}\}\mathbf{h}(\mathbf{x}_i)\mathbf{h}(\mathbf{x}_i)^T$ for generalized linear models (Sec. 4) or $\mathbf{X}_i^T\mathbf{U}_i\mathbf{X}_i$ for multinomial logistic models (Sec. 5). More than that, the design problem discussed here is under more general constraints including but not limited to $n_i \le N_i$, $n_i + n_j \le N_{ij}$ (Example 2), $4n_i \ge n_j$ (Example 3), where N_{ij} is a pre-determined upper bound for the ith and jth categories in total.

For unconstrained optimal design problems, many numerical algorithms have been proposed using directional derivatives (Wynn, 1970; Fedorov, 1972; Atkinson et al., 2014; Fedorov and Leonov, 2014). Among them, the lift-one algorithm (Yang, Mandal and Majumdar, 2016; Yang and Mandal, 2015; Yang, Tong and Mandal, 2017; Bu, Majumdar and Yang, 2020) breaks the problem into univariate optimizations, utilizes analytic solutions whenever possible, reduces unnecessary weights to exact zeros, and works the same well for both local D-optimality and EW D-optimality. A comprehensive numerical study by Yang, Mandal and Majumdar (2016, Tbl. 2) shows that the lift-one algorithm is more efficient than many other commonly used optimization techniques and design algorithms for similar purposes. Unfortunately, it does not work in general for our constrained optimal design problems (Sec. 3.1). The sequential number-theoretic optimization (SNTO) algorithm (Fang and Wang, 1994; Gong et al., 1999; Gao, Liu and Tang, 2022) may provide a solution for our constrained optimization problems, which is, however, typically not as efficient as optimization algorithms based on directional derivatives when the objective function is differentiable and unimodal (Fang, Wang and Bentler, 1994).

In this paper, we develop a new algorithm, called the *constrained lift-one algorithm*, to find optimal allocations under fairly general constraints and statistical models. While keeping the high efficiency of the original lift-one algorithm, the proposed algorithm will check the optimality of the converged allocation and utilize linear programming for adjusting the searching direction when needed. We provide theoretical justifications for the optimality of the allocation found by the proposed algorithm (Sec. 3). Our simulation studies with generalized linear models (Sec. 4) and a real data example with multinomial logistic models (Sec. 5) show that uniformly stratified sampler usually works better than SRSWOR and proportionally stratified sampler, and our designer's choice, (locally) D-optimal and EW D-optimal samplers can significantly improve the efficiency further when some information about the regression coefficients is available.

2. Constrained D-Optimal Allocation

In general, we consider an experiment with $m \ge 2$ pre-determined experimental settings or level combinations $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T \in \mathbb{R}^d$ of d covariates.

Suppose we allocate $n_i \geq 0$ subjects to the ith experimental setting $\mathbf{x}_i$, and the responses are independent and follow a parametric model $M(\mathbf{x}_i; \boldsymbol{\theta})$ with unknown parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$, $p \geq 2$. Under regularity conditions, the Fisher information matrix of the sample can be written as $\mathbf{F} = \sum_{i=1}^m n_i \mathbf{F}_i \in \mathbb{R}^{p \times p}$, where $\mathbf{F}_i$ corresponds to the Fisher information at $\mathbf{x}_i$ and is a positive semi-definite matrix; see, for example, Section 1.5 in Fedorov and Leonov (2014) and references therein.

In design theory, $\mathbf{n} = (n_1, \ldots, n_m)^T$ is called an *exact* allocation of $n = \sum_{i=1}^m n_i$ experimental units, while $\mathbf{w} = (w_1, \ldots, w_m)^T = (n_1/n, \ldots, n_m/n)^T$ is called an *approximate* allocation, which is easier to be dealt with theoretically. The constrained D-optimal design problem considered in this paper is to find the approximate allocation $\mathbf{w}$, which maximizes $|\mathbf{F}|$, the determinant of $\mathbf{F}$, on a collection of feasible approximate allocations $S \subset S_0 := \{(w_1, \ldots, w_m)^T \in \mathbb{R}^m \mid w_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^m w_i = 1\}$. We assume that S is either a closed convex set itself or a finite (overlapped or disjoint) union of closed convex sets. If $S = \cup_{k=1}^K S_k$, where S_k's are all closed convex subsets of S_0, we can always find an optimal allocation for each S_k and then pick up the best one among the optimal allocations. Therefore, in theory, we only need to solve the case when S itself is closed and convex.

Example 1. Consider a paid research study with $N = 500$ eligible volunteers. Suppose $d = 2$ covariates, gender ($x_{i1} = 0$ for female and 1 for male) and age group ($x_{i2} = 0$ for 18–25, 1 for 26–64, and 2 for 65 or above), are important factors. There are $m = 6$ categories with $\mathbf{x}_i = (x_{i1}, x_{i2})^T$ corresponds to $(0,0), (0,1), (0,2), (1,0), (1,1), (1,2)$, respectively. Suppose the numbers of volunteers N_i in the ith category are $50, 40, 10, 200, 150, 50$, respectively. The budget can only support up to $n = 200$ participants who will be under the same treatment. Their responses that will be recorded are binary, 0 for no effect, and 1 for effective. The goal is to study how the effective rate changes along with gender and age group. The collection of feasible approximate allocations is $S = \{(w_1, \ldots, w_6)^T \in S_0 \mid nw_i \leq N_i, i = 1, \ldots, 6\}$, which is closed and convex.

Example 2. Chuang-Stein and Agresti (1997, Tbl. V) provided a dataset of $N = 802$ trauma patients, stratified according to the trauma severity at the time of study entry with 392 mild and 410 moderate/severe patients enrolled. The study involved four treatment groups determined by `dose` level, $x_{i2} = 1$ (`Placebo`), 2 (`Low dose`), 3 (`Medium dose`), and 4 (`High dose`). Combining with `severity` grade ($x_{i1} = 0$ for `mild` or 1 for `moderate/severe`), there are $m = 8$ distinct experimental settings with $(x_{i1}, x_{i2}) = (0,1), (0,2), (0,3), (0,4), (1,1), (1,2), (1,3), (1,4)$, respectively. The responses belong to five ordered categories, (1) `Death`, (2) `Vegetative state`, (3) `Major disability`, (4) `Minor disability` and (5) `Good recovery`, known as the Glasgow Outcome Scale (Jennett and Bond, 1975). Suppose due to a limited budget, only $n = 600$

participants could be supported, the collection of feasible approximate allocations is $S = \{(w_1, \ldots, w_8)^T \in S_0 \mid n(w_1 + w_2 + w_3 + w_4) \le 392, n(w_5 + w_6 + w_7 + w_8) \le 410\}$, which is closed and convex.

In this paper, we adopt the D-optimality, which is to maximize the objective function $f(\mathbf{w}) = |\sum_{i=1}^{m} w_i \mathbf{F}_i|$, $\mathbf{w} \in S$. To avoid trivial cases, we assume that $f(\mathbf{w}) > 0$ for some $\mathbf{w} \in S$. For statistical models under our consideration, such as typical generalized linear models (Sec. 4) and multinomial logistic models (Sec. 5), $\text{rank}(\mathbf{F}_i) < p$ for each $\mathbf{x}_i \in \mathcal{X}$, the collection of all feasible design points, known as the *design space*. Although positive definite $\mathbf{F}_i$ exists for some special multinomial logistic models, it is uncommon and out of the scope of this paper.

Lemma 1. *Suppose* $\text{rank}(\mathbf{F}_i) < p$ *for each* i. *If* $f(\mathbf{w}) > 0$ *for some* $\mathbf{w} = (w_1, \ldots, w_m)^T \in S$, *then* $0 \le w_i < 1$ *for each* i.

Theorem 1. *Suppose* $S \subseteq S_0$ *is closed and* $f(\mathbf{w}) > 0$ *for some* $\mathbf{w} \in S$. *Then* $f(\mathbf{w})$ *is an order-*p *homogeneous polynomial of* $w_1, \ldots, w_m$, *and a D-optimal allocation* $\mathbf{w}_*$ *that maximizes* $f(\mathbf{w})$ *on* S *must exist.*

Lemma 1 and Theorem 1 confirm the existence of the constrained D-optimal allocation. Their proofs, as well as proofs of all other lemmas and theorems, are relegated to Section S6 in the Supplementary Material.

For nonlinear models (Fedorov and Leonov, 2014), generalized linear models (Yang and Mandal, 2015), or multinomial logistic models (Bu, Majumdar and Yang, 2020), $\mathbf{F}$ depends on the unknown parameters $\boldsymbol{\theta}$. We need an assumed $\boldsymbol{\theta}$ to obtain a D-optimal allocation, known as a *locally* D-optimal allocation (Chernoff, 1953). When the investigators only have a rough idea about the parameters, with an assumed prior distribution on $\boldsymbol{\Theta}$, the parameter space, Bayesian D-optimality (Chaloner and Verdinelli, 1995) maximizes $E(\log|\mathbf{F}|)$ and provides a more robust allocation. To overcome its computational intensity, an alternative solution, the EW D-optimality (Atkinson, Donev and Tobias, 2007; Yang, Mandal and Majumdar, 2016), which maximizes $\log|E(\mathbf{F})|$ or $|E(\mathbf{F})|$, was recommended by Yang, Mandal and Majumdar (2016), Yang, Tong and Mandal (2017), and Bu, Majumdar and Yang (2020) for various models. In this paper, we focus on local D-optimality and EW D-optimality.

3. Constrained Lift-One Algorithm

For readers' reference, we provide the original lift-one algorithm for general parametric models as Algorithm 3 in Section S1 of the Supplementary Material. As mentioned in the Introduction section, the original lift-one algorithm does not fit our needs for constrained optimal allocation. We provide such an example (Example 3) in Section 3.1.

3.1. An illustrative example for the lift-one algorithm

In this section, we provide an example such that the allocation found by the original lift-one algorithm is not D-optimal under constraints.

Example 3. Consider an experiment with the logistic regression model $g(\mu_i) = \log\{\mu_i/(1-\mu_i)\} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$ with $\mu_i = E(Y_i \mid \mathbf{x}_i)$ and $\mathbf{x}_i = (x_{i1}, x_{i2})^T \in \{(-1,-1),(-1,+1),(+1,-1)\}$. In this case, $f(\mathbf{w}) = w_1 w_2 w_3$ up to a constant $C > 0$ (see Sec. 4 for more details).

When there is no constraint, as a direct conclusion of the inequality of arithmetic and geometric means, $f(\mathbf{w})$ attains its global maximum at $\mathbf{w}_o = (1/3, 1/3, 1/3)^T \in S_0$ (see Fig. 1 for a 2D display).

Suppose we consider a constrained D-optimal design problem with $S = \{(w_1, w_2, w_3)^T \in S_0 \mid w_1 \le 1/6, w_3 \ge 8/15, 4w_1 \ge w_3\}$, which is a triangle with vertices $\mathbf{w}_a = (1/6, 1/6, 2/3)^T$, $\mathbf{w}_b = (2/15, 1/3, 8/15)^T$ and $\mathbf{w}_d = (1/6, 3/10, 8/15)^T$ (see the shaded region in Fig. 1).

For illustrative purposes, we let the original lift-one algorithm (Alg. 3 in the Supplementary Material) start with $\mathbf{w}_a \in S$. With the order $\{1,3,2\}$ of i, we follow Steps 3°–5° of Algorithm 3 with the ranges of z adjusted according to S (see also Steps 3°–5° of Alg. 1 in Sec. 3.2). At $i = 1$, $f_1(z) = (4/25)z(1-z)^2$ with $z \in \{1/6\}$, which leads to $\mathbf{w}_*^{(1)} = \mathbf{w}_1(1/6) = \mathbf{w}_a$; at $i = 3$, $f_3(z) = (1/4)z(1-z)^2$ with $z \in \{2/3\}$, which leads to $\mathbf{w}_*^{(3)} = \mathbf{w}_3(2/3) = \mathbf{w}_a$; and at $i = 2$, $f_2(z) = (4/25)z(1-z)^2$ with $z \in [1/6, 1/3]$, which is maximized at $z_* = 1/3$ and leads to $\mathbf{w}_*^{(2)} = \mathbf{w}_2(z_*) = \mathbf{w}_b$. That is, after the first round of iterations, $\mathbf{w}_a$ is updated by $\mathbf{w}_b$.

We continue the lift-one iterations with $\mathbf{w}_b$. At $\mathbf{w}_b$, $f_1(z) = (40/169)z(1-z)^2$ with $z \in \{2/15\}$, which leads to $\mathbf{w}_*^{(1)} = \mathbf{w}_1(2/15) = \mathbf{w}_b$; $f_3(z) = (10/49)z(1-z)^2$ with $z \in \{8/15\}$, which leads to $\mathbf{w}_*^{(3)} = \mathbf{w}_3(8/15) = \mathbf{w}_b$; and $f_2(z) = (4/25)z(1-z)^2$ with $z \in [1/6, 1/3]$, which is maximized at $z_* = 1/3$ and leads to $\mathbf{w}_*^{(2)} = \mathbf{w}_2(1/3) = \mathbf{w}_b$. That is, the lift-one algorithm converges at $\mathbf{w}_b$. However, instead of $\mathbf{w}_b$, $\mathbf{w}_d$ is the D-optimal allocation in S (Example 4).

3.2. New algorithm for constrained D-optimal allocations

To find D-optimal allocations under constraints, we develop a new algorithm, called the *constrained lift-one algorithm*, for finding D-optimal allocations in a closed and convex S. If S itself is not convex but a finite union $\cup_{k=1}^K S_k$ of closed and convex S_k's, the proposed algorithm can be applied to each S_k and find the D-optimal allocation $\mathbf{w}_k$ in S_k. Then the D-optimal allocation in S is simply the best one among $\mathbf{w}_k$'s.

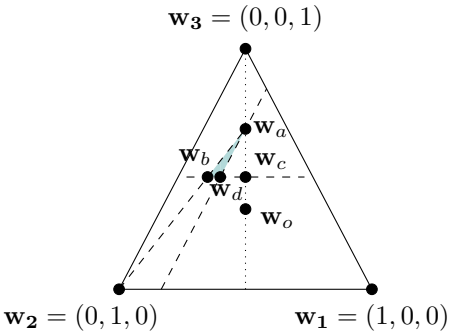

Figure 1. 2D Display of Example 3.

Algorithm 1 Constrained lift-one algorithm under a general setup.

1° Start with an arbitrary allocation $\mathbf{w}_a = (w_1, \ldots, w_m)^T \in S$ satisfying $f(\mathbf{w}_a) > 0$ and $0 \le w_i < 1$, $i = 1, \ldots, m$.

2° Set up a random order of i going through $\{1, 2, \ldots, m\}$. For each i, do steps 3°–5°.

3° For $z \in [0, 1]$, let $\mathbf{w}_i(z) = ((1-z)/(1-w_i)w_1, \ldots, (1-z)/(1-w_i)w_{i-1}, z, (1-z)/(1-w_i)w_{i+1}, \ldots, (1-z)/(1-w_i)w_m)^T$ and $f_i(z) = f\{\mathbf{w}_i(z)\}$. Determine $0 \le r_{i1} \le r_{i2} \le 1$, such that, $\mathbf{w}_i(z) \in S$ if and only if $z \in [r_{i1}, r_{i2}]$.

4° Use an analytic solution or the quasi-Newton algorithm to find z_* maximizing $f_i(z)$ with $z \in [r_{i1}, r_{i2}]$. Define $\mathbf{w}_*^{(i)} = \mathbf{w}_i(z_*)$. Note that $f(\mathbf{w}_*^{(i)}) = f_i(z_*)$.

5° If $f(\mathbf{w}_*^{(i)}) > f(\mathbf{w}_a)$, replace $\mathbf{w}_a$ with $\mathbf{w}_*^{(i)}$, and $f(\mathbf{w}_a)$ with $f(\mathbf{w}_*^{(i)})$.

6° Repeat Steps 2°–5° until convergence, that is, $f(\mathbf{w}_*^{(i)}) \le f(\mathbf{w}_a)$ for each i. Denote $\mathbf{w}_* = (w_1^*, \ldots, w_m^*)^T$ as the converged allocation.

7° Calculate $f_i'(w_i^*)$ for each i. If $f_i'(w_i^*) \le 0$ for all i, then go to Step 10°. Otherwise, go to Step 8°.

8° Find $\mathbf{w}_o \in \operatorname{argmax}_{\mathbf{w} \in S} g(\mathbf{w})$, where $g(\mathbf{w}) = \sum_{i=1}^m w_i(1 - w_i^*) f_i'(w_i^*)$ is a linear function of $\mathbf{w} = (w_1, \ldots, w_m)^T$. If $\mathbf{w}_o$ is not unique, choose any of them. If $g(\mathbf{w}_o) \le 0$, go to Step 10°. Otherwise, go to Step 9°.

9° Use an analytic solution or the quasi-Newton algorithm to find α_* maximizing $h(\alpha) = f\{(1-\alpha)\mathbf{w}_* + \alpha\mathbf{w}_o\}$ with $\alpha \in [0, 1]$ (Thm. 6). Let $\mathbf{w}_a = (1-\alpha_*)\mathbf{w}_* + \alpha_*\mathbf{w}_o$ and go back to Step 2°.

10° Report $\mathbf{w}_*$ as the D-optimal allocation.

Compared with the original lift-one algorithm, Steps 1°–6° in Algorithm 1 are essentially the same except for the intervals $[r_{i1}, r_{i2}]$ in Step 3° due to constraints. Steps 7°–9° in Algorithm 1 are new. Since the lift-one algorithm utilizes the directional derivatives, the searches for optimal allocations are restricted to

the directions between the current allocation and the vertices of S_0. It works for unconstrained optimal design problems but not for constrained ones, since the optimal allocation may not be covered by the directions under constraints (Example 3). In this case, Steps 7° and 8° check whether the current allocation is D-optimal. If not, we use Step 9° to adjust the starting allocation and return searching. Theoretical justifications and more details about Steps 7°–9° can be found in Sections 3.3–3.5.

To find r_{i1} and r_{i2} in Step 3° of Algorithm 1 in general, we suggest the following procedure: (1) Suppose $0 \le w_i < 1$, we start with the interval $I_0 = [0, 1]$ and the line segment $L_0 = \{\mathbf{w}_i(z) \mid z \in I_0\} \subset S_0$; (2) since S is closed and convex, then $L_0 \cap S$ must be closed and convex as well and remain a line segment in the form of $L_i = \{\mathbf{w}_i(z) \mid z \in I_i\}$ with a closed interval $I_i \subseteq [0, 1]$. Then $[r_{i1}, r_{i2}] = I_i$.

A special case is that, as all the examples provided in this paper, S can be written in the form of $\{\mathbf{w} \in S_0 \mid \mathbf{a}_\lambda^T \mathbf{w} \le b_\lambda, \lambda \in \Lambda\}$ with $\mathbf{a}_\lambda \in \mathbb{R}^m$ and $b_\lambda \in \mathbb{R}$. For each $\lambda \in \Lambda$, $\mathbf{a}_\lambda^T \mathbf{w}_i(z) \le b_\lambda$ with $z \in [0, 1]$ implies $\{\mathbf{w}_i(z) \mid z \in I_\lambda\}$ with some interval $I_\lambda \subseteq [0, 1]$. Then $[r_{i1}, r_{i2}] = \cap_{\lambda\in\Lambda} I_\lambda$, which is nonempty (Examples 9 and 10 in the Supplementary Material).

Example 4. Example 3 is considered here again. Recall that $\mathbf{w}_b = (2/15, 1/3, 8/15)^T$ is reported as the converged allocation in Step 6° of Algorithm 3 (or Alg. 1). To check the conditions in Step 7° of Algorithm 1 for $\mathbf{w}_b$, we obtain $f_1'(2/15) = 8/65 > 0$, $f_2'(1/3) = 0$, and $f_3'(8/15) = -2/35 < 0$. We then go to Step 8° with $g(\mathbf{w}) = (2/75)(4w_1 - w_3)$. Since S is the convex hull of its vertex set $\{\mathbf{w}_a, \mathbf{w}_b, \mathbf{w}_d\}$, it can be verified that (Sec. 3.4, Thm. 4) $\mathbf{w}_d = (1/6, 3/10, 8/15)^T$ maximizes $g(\mathbf{w})$, $\mathbf{w} \in S$. Since $g(\mathbf{w}_d) = (4/1125) > 0$, we go to Step 9° and define $h(\alpha) = f\{(1-\alpha)\mathbf{w}_b + \alpha\mathbf{w}_d\} = (2/15^3)(4+\alpha)(10-\alpha)$. Since $h'(\alpha) = (4/15^3)(3-\alpha) > 0$ for all $\alpha \in [0, 1]$, then $\alpha_* = \text{argmax}_{\alpha\in[0,1]} h(\alpha) = 1$. We let $\mathbf{w}_a^{(1)} = (1-\alpha_*)\mathbf{w}_b + \alpha_*\mathbf{w}_d = \mathbf{w}_d$ and go back to Step 2° of Algorithm 1.

First of all, $\mathbf{w}_d$ is a converged allocation in Step 6°. Actually, $f_1(z) = (144/625)z(1-z)^2$ with $z \in [4/29, 1/6]$ and is maximized at $z_* = (1/6)$; $f_2(z) = (80/441)z(1-z)^2$ with $z \in \{3/10\}$ and thus is maximized at $z_* = 3/10$; and $f_3(z) = (45/196)z(1-z)^2$ with $z \in [8/15, 10/17]$ and is maximized at $z_* = 8/15$. Secondly, since $f_1'(1/6) = 12/125 > 0$, $f_2'(3/10) = 4/315 > 0$, and $f_3'(8/15) = -9/140 < 0$, we go to Step 8°. Since $g(\mathbf{w}) = (2/25)(w_1 + w_2/9 - (3/8)w_3)$ is maximized at $\mathbf{w}_d$ (see Thm. 4 again) and $g(\mathbf{w}_d) = 0$, we go to Step 10° and report $\mathbf{w}_d$ as the D-optimal allocation.

Given an approximate allocation $\mathbf{w} = (w_1, \ldots, w_m)^T \in S$, if all additional constraints for S take the form of $\sum_{i=1}^m a_i w_i \le c$ with $a_i \ge 0$ and $c > 0$ such as in Examples 1 and 2, we develop the following constrained round-off algorithm to obtain an exact allocation $\mathbf{n} = (n_1, \ldots, n_m)^T$ satisfying $\mathbf{n}/n \in S$ and $\sum_{i=1}^m n_i \le n$.

The allocation obtained by Algorithm 1 with $f(\mathbf{w}) = |\sum_{i=1}^m w_i \mathbf{F}_i|$ is known as a *locally D-optimal* allocation since it may require assumed values of $\boldsymbol{\theta}$. With a

Algorithm 2 Constrained round-off algorithm for obtaining a feasible exact allocation.

1° First let $n_i = \lfloor nw_i \rfloor$, the largest integer no more than nw_i, $i = 1, \ldots, m$, and $k = n - \sum_{i=1}^{m} n_i$. Denote $I = \{i \in \{1, \ldots, m\} \mid w_i > 0, (n_1, \ldots, n_{i-1}, n_i + 1, n_{i+1}, \ldots, n_m)/n \in S\}$.

2° While $k > 0$ and $I \neq \emptyset$, do

2.1 For $i \in I$, calculate $d_i = f(n_1, \ldots, n_{i-1}, n_i + 1, n_{i+1}, \ldots, n_m)$.

2.2 Pick up any $i_* \in \operatorname{argmax}_{i \in I} d_i$.

2.3 Let $n_{i_*} \leftarrow n_{i_*} + 1$ and $k \leftarrow k - 1$.

2.4 If $(n_1, \ldots, n_{i_*-1}, n_{i_*} + 1, n_{i_*+1}, \ldots, n_m)/n \notin S$, then $I \leftarrow I \setminus \{i_*\}$.

3° Output $\mathbf{n} = (n_1, \ldots, n_m)^T$.

specified prior distribution $h(\boldsymbol{\theta})$ on the parameter space $\boldsymbol{\Theta}$, we may replace $f(\mathbf{w})$ with $f_{\mathrm{EW}}(\mathbf{w}) = |\sum_{i=1}^{m} w_i E(\mathbf{F}_i)|$ and the obtained allocation by Algorithm 1 is called an *EW D-optimal* allocation (Sec. 2).

3.3. D-optimality of Algorithm 1

In this section, we show that the allocation reported by Algorithm 1 is D-optimal. Throughout this section, we assume that S is closed and convex.

Lemma 2. *Suppose* $f(\mathbf{w}_a) > 0$ *for some* $\mathbf{w}_a = (w_1, \ldots, w_m)^T \in S$ *with* $0 \leq w_i < 1$, $i = 1, \ldots, m$. *Let* $f_i(z) = f\{\mathbf{w}_i(z)\}$ *as defined in Algorithm* 1. *Then* $\log f_i(z)$ *is a concave function on* $[r_{i1}, r_{i2}]$. *Furthermore, suppose* z_* *maximizes* $f_i(z)$ *on* $[r_{i1}, r_{i2}]$. *Then* (1) *if* $z_* = r_{i1} < r_{i2}$, *then* $f_i'(z_*) \leq 0$; (2) *if* $z_* = r_{i2} > r_{i1}$, *then* $f_i'(z_*) \geq 0$; *and* (3) *if* $z_* \in (r_{i1}, r_{i2})$, *then* $f_i'(z_*) = 0$.

Theorem 2. *Suppose* $f(\mathbf{w}) > 0$ *for some* $\mathbf{w} \in S$. *Let* $\mathbf{w}_* = (w_1^*, \ldots, w_m^*)^T \in S$ *be a converged allocation in Step* 6° *of Algorithm* 1 *with* $0 \leq w_i^* < 1$, $i = 1, \ldots, m$. *If* $f_i'(w_i^*) \leq 0$ *for each* i, *then* $\mathbf{w}_*$ *must be D-optimal in* S.

Using Lemma 2, Theorem 2 and the following corollary, the D-optimality of the converged allocation $\mathbf{w}_*$ under some conditions can be easily justified.

Corollary 1. *Suppose* $\operatorname{rank}(\mathbf{F}_i) < p$ *for each* i *and* $f(\mathbf{w}) > 0$ *for some* $\mathbf{w} \in S$. *Let* $\mathbf{w}_* \in S$ *be a converged allocation in Step* 6° *of Algorithm* 1. *If* $w_i^* < r_{i2}$ *for each* i, *then* $\mathbf{w}_*$ *must be D-optimal in* S.

Remark 1. If $S = S_0$, then $[r_{i1}, r_{i2}]$ in Step 3° of Algorithm 1 is $[0, 1]$ for each i. Let $\mathbf{w}_* = (w_1^*, \ldots, w_m^*)^T \in S_0$ be a converged allocation in Step 6° of Algorithm 1. If $\operatorname{rank}(\mathbf{F}_i) < p$ for each i and $f(\mathbf{w}) > 0$ for some $\mathbf{w} \in S$, then $w_i^* < r_{i2} = 1$ for each i. According to Corollary 1, $\mathbf{w}_*$ must be D-optimal in S_0. That is, the original lift-one algorithm (Alg. 3 in the Supplementary Material) still works for the case $S = S_0$.

Theorem 3. *Suppose* $\text{rank}(\mathbf{F}_i) < p$ *for each* i *and* $f(\mathbf{w}) > 0$ *for some* $\mathbf{w} \in S$. *Then* $\mathbf{w}_*$ *reported in Step* 10° *of Algorithm* 1 *is D-optimal in* S.

3.4. Maximization of $g(\mathbf{w})$ in Step 8° of Algorithm 1

In this section, we provide the solutions maximizing $g(\mathbf{w}) = \sum_{i=1}^m w_i(1 - w_i^*)f_i'(w_i^*)$ with $\mathbf{w} = (w_1, \ldots, w_m)^T \in S$, where $S \subseteq S_0$ is closed and convex. By letting $a_i = (1 - w_i^*)f_i'(w_i^*)$ and $\mathbf{a} = (a_1, \ldots, a_m)^T$, $g(\mathbf{w}) = \mathbf{a}^T\mathbf{w}$ is a linear function of $\mathbf{w} \in S$.

For many applications including all the examples considered in this paper, S is determined by linear conditions or constraints and the maximization of $g(\mathbf{w})$ can be written as:

$$\begin{aligned} &\text{Max } \mathbf{a}^T\mathbf{w} \\ &\text{subject to } \mathbf{G}\mathbf{w} \preceq \mathbf{h},\ \mathbf{A}\mathbf{w} = \mathbf{b}, \end{aligned} \tag{3.1}$$

where $\mathbf{G} \in \mathbb{R}^{r\times m}$, $\mathbf{h} \in \mathbb{R}^r$, $\mathbf{A} \in \mathbb{R}^{s\times m}$, $\mathbf{b} \in \mathbb{R}^s$ are known matrices or vectors, and "$\preceq$" is componentwise "$\le$". It is known as a *linear program* (LP) problem (Boyd and Vandenberghe, 2004, Sec. 4.3) and can be efficiently solved by using, for example, `R` function `lp` in Package `lpSolve`.

For general cases, $S \subseteq S_0$ is closed and convex. Since $S_0 \subset \mathbb{R}^m$ is bounded, S is bounded and thus compact. According to Theorem 5.6 in Lay (1982), S is the convex hull of its profile E, the set consisting of all extreme points of S. Since $g(\mathbf{w})$ is linear on S which depends on $\mathbf{w}_*$, according to Theorem 5.7 in Lay (1982), there exists a $\mathbf{w}_o \in E$ such that

$$\mathbf{w}_o \in \operatorname*{argmax}_{\mathbf{w}\in E} g(\mathbf{w}) = \operatorname*{argmax}_{\mathbf{w}\in S} g(\mathbf{w}).$$

In other words, we only need to search $\mathbf{w}_o$ among the profile E of S, which is only a subset of the boundary of S. According to the proof of Theorem 2, if we can find a $\mathbf{w}_c \in S$ such that $f(\mathbf{w}_c) > f(\mathbf{w}_*)$, then $g(\mathbf{w}_c) > 0$. Since $g(\mathbf{w}_c)$ is the directional derivative of $f(\mathbf{w})$ along $\mathbf{w}_c - \mathbf{w}_*$, $g(\mathbf{w}_c) > 0$ implies that there exists at least one point along the direction with higher $f(\mathbf{w})$ value, which is not necessary to be $f(\mathbf{w}_c)$. Following the proof of Theorem 5.7 in Lay (1982), we obtain the following convenient results for special cases:

Theorem 4. *If there exists a* $V = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\} \subset S$, *such that,* S *can be rewritten as* $\{b_1\mathbf{w}_1 + \cdots + b_k\mathbf{w}_k \mid b_1 \ge 0, \ldots, b_k \ge 0, \sum_{i=1}^k b_i = 1\}$, *then* $\mathbf{w}_o \in \operatorname{argmax}_{\mathbf{w}\in V} g(\mathbf{w})$ *maximizes* $g(\mathbf{w})$ *on* S.

In other words, if S is the convex hull of a finite set V (such an S is called a *polytope* or *convex polytope* in the literature; see, for example, Def. 2.24 in Lay (1982)), we only need to search $\mathbf{w}_o$ among the finite set V. Note that the V, known as the *vertex set*, may not be unique and may not be the profile of S.

Given $a_i = (1-w_i^*)f_i'(w_i^*)$, $i = 1, \ldots, m$, we call $r_1, \ldots, r_m$ the *ranks* of them, if $\{r_1, \ldots, r_m\} = \{1, \ldots, m\}$ and $a_{r_1} \geq a_{r_2} \geq \cdots \geq a_{r_m}$. For S taking the form as in Example 1, we have an analytic solution for $\mathbf{w}_o$:

Theorem 5. *Suppose* $S = \{\mathbf{w} \in S_0 \mid w_i \leq c_i, i = 1, \ldots, m\}$ *with* $0 < c_i \leq 1$, $i = 1, \ldots, m$ *and* $\sum_{i=1}^m c_i \geq 1$. *Then a* $\mathbf{w}_o = (w_1^o, \ldots, w_m^o)^T$ *maximizing* $g(\mathbf{w}) = \sum_{i=1}^m a_i w_i$ *can be obtained as follows: (i) if* $\sum_{i=1}^m c_i = 1$, $\mathbf{w}_o = (c_1, \ldots, c_m)^T$; *(ii) if* $\sum_{i=1}^m c_i > 1$, *then* $w_i^o = c_i$ *if* $i \in \{r_1, \ldots, r_k\}$; $1 - \sum_{l=1}^k c_{r_l}$ *if* $i = r_{k+1}$; *and* 0 *otherwise, where* $r_1, \ldots, r_m$ *are the ranks of* $a_1, \ldots, a_m$ *and* $k \in \{1, \ldots, m-1\}$ *satisfying* $\sum_{l=1}^k c_{r_l} \leq 1 < \sum_{l=1}^{k+1} c_{r_l}$.

Example 5. Example 3 is considered here again. In this case, $S = \{(w_1, w_2, w_3)^T \in S_0 \mid w_1 \leq 1/6, w_3 \geq 8/15, 4w_1 \geq w_3\}$ with its vertex set $V = \{\mathbf{w}_a, \mathbf{w}_b, \mathbf{w}_d\}$ (Fig. 1). As mentioned in Example 4, $g(\mathbf{w})$ defined with both $\mathbf{w}_b$ and $\mathbf{w}_d$ is maximized in S at $\mathbf{w}_d$, one of the three vertices.

3.5. Maximization of $h(\alpha)$ in Step 9° of Algorithm 1

Suppose $\mathbf{w}_* = (w_1^*, \ldots, w_m^*)^T \in S$, $0 \leq w_i^* < 1$ for all i, is a converged allocation in Step 6° of Algorithm 1, and $\mathbf{w}_o = (w_1^o, \ldots, w_m^o)^T \in S$ is obtained in Step 8° with $g(\mathbf{w}_o) > 0$. We provide the following results to find α_* maximizing $h(\alpha) = f\{(1-\alpha)\mathbf{w}_* + \alpha\mathbf{w}_o\}$ in Step 9°.

Lemma 3. *The function* $h(\alpha)$ *in Step* 9° *of Algorithm* 1 *can be written as*

$$h(\alpha) = c_0 + c_1\alpha + \cdots + c_{p-1}\alpha^{p-1} + c_p\alpha^p, \tag{3.2}$$

where $c_0 = f(\mathbf{w}_*)$, $(c_1, \ldots, c_p)^T = \mathbf{B}_p^{-1}(h(1/p) - c_0, \ldots, h((p-1)/p) - c_0, h(1) - c_0)^T$, *and* $\mathbf{B}_p$ *is a* $p \times p$ *matrix with its* (s,t)*th entry* $(s/p)^t$.

Based on Lemma 3, we can determine the coefficients of $h(\alpha)$ with $h(1/p), \ldots, h((p-1)/p)$, and $h(1) = f(\mathbf{w}_o)$, and then calculate $h'(\alpha)$ by

$$h'(\alpha) = c_1 + 2c_2\alpha + \cdots + (p-1)c_{p-1}\alpha^{p-2} + pc_p\alpha^{p-1}. \tag{3.3}$$

Theorem 6. *Suppose* $f(\mathbf{w}_*) > 0$, $g(\mathbf{w}_o) = \sum_{i=1}^m w_i^o(1-w_i^*)f_i'(w_i^*) > 0$, $h(\alpha) = f\{(1-\alpha)\mathbf{w}_* + \alpha\mathbf{w}_o\}$, *and* α_* *maximizes* $h(\alpha)$ *on* $[0,1]$ *as defined in Step* 9° *of Algorithm* 1. *Then (i)* $h(\alpha) > 0$ *for all* $\alpha \in [0,1)$; *(ii)* $h'(0) > 0$; *(iii) if* $h(1) > 0$ *and* $h'(1) \geq 0$, *then* $\alpha_* = 1$; *(iv) if* $h(1) > 0$ *and* $h'(1) < 0$, *or* $h(1) = 0$, *then there exists a unique* $\alpha_0 \in (0,1)$ *such that* $h'(\alpha_0) = 0$, *which implies* $\alpha_* = \alpha_0$. *By combining (iii) and (iv), we know that* α_* *always exists and is unique.*

Based on Theorem 6 and Equation (3.3), if $h'(1) < 0$, one may use, for example, the `R` function `uniroot`, to find $\alpha_* \in (0,1)$ numerically. To avoid numerical errors, we need to double check $\alpha_* \neq 1$ when $f(\mathbf{w}_o) < f(\mathbf{w}_*)$.

In Step 9° of Algorithm 1, we let $\mathbf{w}_a = (1-\alpha_*)\mathbf{w}_* + \alpha_*\mathbf{w}_o$. According to Theorem 6 (ii), $h'(0) > 0$, which implies $f(\mathbf{w}_a) = h(\alpha_*) > f(\mathbf{w}_*)$. Nevertheless, it does not mean that $\mathbf{w}_a$ is D-optimal already.

4. D-Optimal Samplers for Generalized Linear Models

In this section, we utilize local and EW D-optimal samplers for univariate responses, such as in Example 1. Recall that we assign n_i participants to the ith category. We let Y_{ij} stand for the univariate response of the jth participant of the ith category. Generalized linear models (McCullagh and Nelder, 1989; Dobson and Barnett, 2018)

$$E(Y_{ij} \mid \mathbf{x}_i) = \mu_i \text{ and } g(\mu_i) = \eta_i = \mathbf{h}(\mathbf{x}_i)^T\boldsymbol{\theta} \tag{4.1}$$

have been widely used, where $i = 1, \ldots, m$; $j = 1, \ldots, n_i$; g is a given (*link*) function; η_i is known as a linear predictor; $\mathbf{h}(\mathbf{x}_i) = (h_1(\mathbf{x}_i), \ldots, h_p(\mathbf{x}_i))^T$ are p given predictor functions; and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ are the regression coefficients. Commonly used generalized linear models (GLM) (Tbl. 5 in the Supplementary Material) cover Gaussian response (that is, linear models), binary response (Bernoulli, such as Example 1), count response (Poisson), and real positive response (Gamma, Inverse Gaussian).

Assuming that Y_{ij}'s are independent, the Fisher information matrix (Yang and Mandal, 2015)

$$\mathbf{F}(\mathbf{w}) = n\sum_{i=1}^{m} w_i\mathbf{F}_i = n\sum_{i=1}^{m} w_i\nu_i\mathbf{h}(\mathbf{x}_i)\mathbf{h}(\mathbf{x}_i)^T = n\mathbf{X}^T\mathbf{W}\mathbf{X},$$

where $\mathbf{w} = (w_1, \ldots, w_m)^T$, $w_i = n_i/n$, $\mathbf{X} = (\mathbf{h}(\mathbf{x}_1), \ldots, \mathbf{h}(\mathbf{x}_m))^T$ is an $m \times p$ matrix, $\mathbf{W} = \text{diag}\{w_1\nu_1, \ldots, w_m\nu_m\}$, and $\nu_i = \nu(\eta_i) = (\partial\mu_i/\partial\eta_i)^2/\text{Var}(Y_{ij})$, $i = 1, \ldots, m$. We provide examples of $\nu(\eta_i)$ for commonly used GLMs in Table 5 of the Supplementary Material. For GLMs, $f(\mathbf{w}) = |\mathbf{X}^T\mathbf{W}\mathbf{X}|$ and $f_{\text{EW}}(\mathbf{w}) = |\mathbf{X}^T E(\mathbf{W})\mathbf{X}|$ (Sec. 2).

According to Lemma 4.1 in Yang and Mandal (2015), $f_i(z) = az(1-z)^{p-1} + b(1-z)^p$, where $b = f_i(0)$, $a = \{f(\mathbf{w}) - b(1-w_i)^p\}/\{w_i(1-w_i)^{p-1}\}$ if $w_i > 0$; and $b = f(\mathbf{w})$, $a = f_i(1/2)2^p - b$ otherwise. In both cases, $a \geq 0$, $b \geq 0$, and $a + b > 0$. To implement Step 7° of Algorithm 1, we need

$$f_i'(z) = \{a - bp + (b-a)pz\}(1-z)^{p-2}. \tag{4.2}$$

Here $m \geq p \geq 2$. Similar to Lemma 4.2 in Yang and Mandal (2015), we provide the following lemma for maximizing $f_i(z)$ with constraints:

Lemma 4. *Denote* $l(x) = ax(1-x)^{p-1} + b(1-x)^p$ *with* $a \geq 0$, $b \geq 0$ *and* $a+b > 0$. *Let* $\delta = (a-bp)/\{(a-b)p\}$ *when* $a \neq b$. *Then*

$$x_* = \begin{cases} \delta, & \text{if } a > bp \text{ and } r_1 \le \delta \le r_2; \\ r_2, & \text{if } a > bp \text{ and } \delta > r_2; \\ r_1, & \text{otherwise.} \end{cases} \tag{4.3}$$

maximizes $l(x)$ *with constraints* $0 \le r_1 \le x \le r_2 \le 1$.

Example 6. Example 1 is considered here again. In this case, $N = 500$ eligible volunteers are available for $m = 6$ categories with frequencies $(N_1, N_2, \ldots, N_6) = (50, 40, 10, 200, 150, 50)$. For illustration purposes, we consider a logistic regression model (GLM with Bernoulli(μ_i) and logit link):

$$\text{logit}\{P(Y_{ij} = 1 \mid x_{i1}, x_{i2})\} = \beta_0 + \beta_1 x_{i1} + \beta_{21} 1_{\{x_{i2}=1\}} + \beta_{22} 1_{\{x_{i2}=2\}}, \tag{4.4}$$

where $i = 1, \ldots, 6$; $j = 1, \ldots, n_i$; and $\text{logit}(\mu) = \log\{\mu/(1-\mu)\}$. Model (4.4) is a main-effects model with gender and age group as factors.

To sample $n = 200$ from $m = 6$ categories or strata, the proportionally stratified allocation is $\mathbf{w}_\text{p} = (0.10, 0.08, 0.02, 0.40, 0.30, 0.10)^T$ or $\mathbf{n}_\text{p} = (20, 16, 4, 80, 60, 20)^T$, while the (constrained) uniformly stratified allocation is $\mathbf{w}_\text{u} = (0.19, 0.19, 0.05, 0.19, 0.19, 0.19)^T$ or $\mathbf{n}_\text{u} = (38, 38, 10, 38, 38, 38)^T$. By implementing Algorithms 1 and 2 in `R` with assumed $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_{21}, \beta_{22})^T = (0, 3, 3, 3)^T$, we obtain the (locally) D-optimal allocation $\mathbf{w}_\text{o} = (0.25, 0.20, 0.05, 0.50, 0, 0)^T$ or $\mathbf{n}_\text{o} = (50, 40, 10, 100, 0, 0)^T$. Compared with $\mathbf{w}_\text{o}$, the relative efficiency of $\mathbf{w}_\text{p}$ is $\{|\mathbf{F}(\mathbf{w}_\text{p})|/|\mathbf{F}(\mathbf{w}_\text{o})|\}^{1/p} = 53.93\%$ with the number of parameters $p = 4$, and the relative efficiency of $\mathbf{w}_\text{u}$ is $\{|\mathbf{F}(\mathbf{w}_\text{u})|/|\mathbf{F}(\mathbf{w}_\text{o})|\}^{1/p} = 78.99\%$. Both are much less efficient than $\mathbf{w}_o$.

We also look into the robustness of our optimal allocations to model misspecification. Assuming that the true link in this study is not the assumed logit link but probit, complementary log-log, or log-log link (Tbl. 5 in the Supplementary Material), we check the relative efficiency of the allocation obtained under logit to the D-optimal allocations based on the true link. Notice that the log-log link shares the same $\mathbf{W}$ matrix as the complementary log-log link. Since $\boldsymbol{\beta} = (0, 3, 3, 3)^T$ satisfies the conditions of Theorem 3.2 in Yang and Mandal (2015), the D-optimal designs are saturated and different links lead to the same D-optimal allocation. Therefore, the relative efficiency of our proposed allocation under logit remains 100% with respect to link misspecifications. In Section S4 of the Supplementary Material, we provide an example with different assumed parameter values, which still have 99% relative efficiencies with link misspecifications. We also provide the results based on the root mean squared errors (RMSE) in Table 6 in the Supplementary Material, which is consistent with the relative efficiency result and confirms the robustness of our proposed allocations.

To compare the accuracy of the estimated regression coefficients based on different samplers, we use the RMSE $(\{\sum_{i \in I}(\hat{\beta}_i - \beta_i)^2/|I|\}^{1/2}$ given an index

Table 1. Average (sd) of RMSE over 100 Simulations under Model (4.4).

Sampler	Average (sd) of RMSE				
	β_0	all except β_0	β_1	β_{21}	β_{22}
Full Data	0.195(0.145)	6.317(4.070)	0.363(0.289)	2.751(5.468)	9.098(7.018)
SRSWOR	0.314(0.216)	9.984(3.226)	0.917(2.543)	8.098(7.885)	12.976(5.245)
Proportionally Stratified	0.412(0.304)	9.496(3.682)	1.016(2.545)	7.311(7.942)	12.469(5.682)
Uniformly Stratified	0.235(0.193)	7.967(4.659)	3.855(6.673)	3.353(6.254)	9.657(7.297)
Locally D-opt	0.202(0.150)	7.103(4.098)	0.485(0.438)	3.890(6.507)	9.883(6.821)
Unif EW D-opt	0.201(0.145)	7.556(4.653)	1.538(3.942)	3.920(6.561)	9.273(7.151)
Normal EW D-opt	0.202(0.147)	7.252(4.407)	1.347(3.664)	3.982(6.687)	9.302(7.150)
Gamma EW D-opt	0.205(0.153)	7.718(4.476)	1.535(4.008)	3.955(6.652)	9.585(7.080)

set I). For illustration purposes, we assume that the true parameters are $(\beta_0, \beta_1, \beta_{21}, \beta_{22}) = (0, 3, 3, 3)$ and run 100 simulations. In each simulation, we generate $N = 500$ independent observations based on Model (4.4) and use SRSWOR, proportionally stratified sampler, uniformly stratified sampler, and D-optimal sampler, respectively, to sample $n = 200$ observations out of 500. We then fit Model (4.4) using the $n = 200$ observations to get the estimated parameters $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_{21}, \hat{\beta}_{22})$. The average and standard deviation (sd) of RMSEs across 100 simulations are listed in Tbl. 1. According to the RMSE with index set $I = \{1, 21, 22\}$ (column "all except β_0" in Tbl. 1), SRSWOR is the least accurate, proportional stratified sampler is a little better, uniformly stratified sampler is much better, and locally D-optimal sampler is the best, which is much closer to the full data estimates. For readers' reference, we also list the RMSEs for individual β_i's.

If we know something about $\boldsymbol{\theta}$ but not their exact values, we recommend EW D-optimal samplers instead (see also Sec. 2). For illustration purposes, we consider three different prior distributions: (i) uniform prior: $\boldsymbol{\theta} \sim \text{Unif}(-2, 2) \times \text{Unif}(-1, 5) \times \text{Unif}(-1, 5) \times \text{Unif}(-1, 5)$; (ii) normal prior: $h(\boldsymbol{\theta}) = \phi(\beta_0/0.5) \times \phi\{(\beta_1 - 2)/0.5\} \times \phi\{(\beta_{21} - 2)/0.5\} \times \phi\{(\beta_{22} - 2)/0.5\}$; (iii) Gamma prior: $\boldsymbol{\theta} \sim N(0, 1) \times \text{Gamma}(1, 2) \times \text{Gamma}(1, 2) \times \text{Gamma}(1, 2)$. The relevant expectations $E\left[\nu\{\mathbf{h}^T(\mathbf{x}_i)\boldsymbol{\theta}\}\right]$ can be numerically computed using, for example, `R` function `hcubature` in package `cubature`. Compared with the locally D-optimal allocation $\mathbf{w}_{\text{o}}$ which samples only from four categories, EW allocations are not so extreme, such as $\mathbf{w}_{\text{uEW}} = (0.240, 0.200, 0.050, 0.211, 0.101, 0.198)^T$ with the uniform prior, $\mathbf{w}_{\text{nEW}} = (0.250, 0.200, 0.050, 0.334, 0, 0.166)^T$ with the normal prior, and $\mathbf{w}_{\text{gEW}} = (0.240, 0.200, 0.050, 0.214, 0.096, 0.200)^T$ with the gamma prior. Compared with $\mathbf{w}_{\text{o}}$, their relative efficiencies are 85.90%, 94.96%, and 86.32%, respectively, which are still much better than SRSWOR, proportionally stratified, and uniformly stratified samplers. In terms of RMSE (Tbl. 1), the conclusions are consistent.

Known as the *uniform* allocation, $\mathbf{w} = (1/m, \ldots, 1/m)^T$ has a special role in optimal design theory, which is recommended for linear models or as a robust design (Yang, Mandal and Majumdar, 2012). In this paper, we introduce

constrained uniform allocations such as $\mathbf{w}_{\mathrm{u}}$ in Example 1. They are D-optimal for saturated cases (that is, $m = p$).

Lemma 5. *Suppose* $\mathbf{w}_* = (w_1^*, \ldots, w_m^*)^T$ *maximizes* $f(\mathbf{w}) = \prod_{i=1}^m w_i$ *under the constraints* $0 \le w_i \le c_i$, $i = 1, \ldots, m$ *and* $\sum_{i=1}^m w_i = 1$, *where* $0 < c_i \le 1$, $i = 1, \ldots, m$ *and* $\sum_{i=1}^m c_i \ge 1$. *Then (i) if* $\min_{1 \le i \le m} c_i \ge 1/m$, *then* $\mathbf{w}_* = (1/m, \ldots, 1/m)^T$; *(ii) if* $0 \le \min_{1 \le i \le m} c_i < 1/m$ *and* $\sum_{i=1}^m c_i > 1$, *then there exist* $1 \le k \le m-1$ *and* $c_{(k)} \le u < c_{(k+1)}$, *such that,* $w_i^* = c_i$ *if* $c_i \le u$, *and* $w_i^* = u$ *if* $c_i > u$, *where* $0 < c_{(1)} \le c_{(2)} \le \cdots \le c_{(m)} \le 1$ *are order statistics of* $c_1, \ldots, c_m$; *(iii) if* $\sum_{i=1}^m c_i = 1$, *then* $w_i^* = c_i$, $i = 1, \ldots, m$.

We call the $\mathbf{w}_*$ described in Lemma 5 a *constrained uniform allocation* and the corresponding sampler a (constrained) *uniformly stratified* sampler.

Theorem 7. *For GLM* (4.1) *with* $m = p$, *if* $S = \{\mathbf{w} \in S_0 \mid w_i \le c_i, i = 1, \ldots, m\}$ *with* $0 < c_i \le 1$, $i = 1, \ldots, m$ *and* $\sum_{i=1}^m c_i \ge 1$, *then the constrained uniform allocation described in Lemma* 5 *is both D-optimal and EW D-optimal.*

Example 7. Example 1 continues here. If we consider another logistic regression model

$$\begin{aligned}\operatorname{logit}\{P(Y_{ij} = 1 \mid x_{i1}, x_{i2})\} = \beta_0 + \beta_1 x_{i1} + \beta_{21} 1_{\{x_{i2}=1\}} + \beta_{22} 1_{\{x_{i2}=2\}} \\ + \beta_{121} x_{i1} 1_{\{x_{i2}=1\}} + \beta_{122} x_{i1} 1_{\{x_{i2}=2\}} \qquad (4.5)\end{aligned}$$

for Example 1, which adds two order-2 interactions to Model (4.4). Then $m = p = 6$. According to Theorem 7, the constrained uniform allocation $\mathbf{w}_{\mathrm{u}} = (0.19, 0.19, 0.05, 0.19, 0.19, 0.19)^T$ is both D-optimal and EW D-optimal. In this case, the uniformly stratified sampler is the same as the D-optimal and EW D-optimal samplers.

To compare the SRSWOR, proportionally stratified, uniformly stratified/locally D-optimal/EW D-optimal samplers, for illustration purposes, we assume that the true parameters are $(\beta_0, \beta_1, \beta_{21}, \beta_{22}, \beta_{121}, \beta_{122}) = (0, -0.1, -0.5, -2, -0.5, -1)$. We run 100 simulations using Model (4.5). In this scenario, the proportionally stratified allocation $\mathbf{w}_{\mathrm{p}}$ and the uniformly stratified allocation $\mathbf{w}_{\mathrm{u}}$ are the same as in Example 1, and the D-optimal allocation $\mathbf{w}_{\mathrm{o}} = \mathbf{w}_{\mathrm{u}}$. The relative efficiencies of $\mathbf{w}_{\mathrm{p}}$ and $\mathbf{w}_{\mathrm{u}}$ compared with $\mathbf{w}_{\mathrm{o}}$ are 73.30% and 100%, respectively. In terms of robustness to model misspecifications, the relative efficiencies with true links as the probit, log-log, and complementary log-log are again 100% due to Theorem 7. The average and standard deviation of the 100 RMSEs are reported in Table 2. Again, the D-optimal sampler (same as the uniformly stratified sampler in this scenario) significantly reduces the RMSEs based on SRSWOR or the proportionally stratified sampler.

Table 2. Average (sd) of RMSE over 100 Simulations under Model (4.5)

Sampler	Average (sd) of RMSEs						
	β_0	all except β_0	β_1	β_{21}	β_{22}	β_{121}	β_{122}
Full Data	0.240(0.206)	3.753(3.752)	0.285(0.230)	0.350(0.263)	4.622(6.406)	0.403(0.319)	5.004(6.343)
SRSWOR	0.340(0.268)	6.230(3.348)	0.386(0.290)	0.552(0.398)	9.005(6.826)	0.613(0.508)	7.184(6.845)
Proportionally Stratified	0.379(0.316)	6.347(3.267)	0.459(0.369)	0.593(0.512)	9.400(6.877)	0.742(0.568)	6.757(6.904)
Uniformly/D-opt/EW D-opt	0.267(0.203)	4.186(3.944)	0.425(0.304)	0.367(0.272)	4.973(6.936)	0.602(0.497)	5.485(6.887)

5. D-optimal Samplers for Multinomial Logit Models

In this section, we utilize D-optimal samplers for categorical responses as in Example 2. For the ith experimental setting $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T$, n_i categorical responses are collected i.i.d. from a discrete distribution with $J \geq 2$ categories, $i = 1, \ldots, m$. The summary statistics $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})^T \sim$ Multinomial$(n_i; \pi_{i1}, \ldots, \pi_{iJ})$, where Y_{ij} is the number of responses of the jth category, π_{ij} is the probability that the response falls into the jth category at $\mathbf{x}_i$. Assuming $\pi_{ij} > 0$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, J$, multinomial logit models have been widely used in the literature (see Bu, Majumdar and Yang (2020) and references therein), including commonly used baseline-category, cumulative, adjacent-categories, and continuation-ratio logit models.

Example 8. Example 2 is considered here again. In this case, the response has $J = 5$ categories, and there are $m = 8$ distinct experimental settings determined by $d = 2$ factors, `severity` ($x_{i1} \in \{0, 1\}$) and `dose` ($x_{i2} \in \{1, 2, 3, 4\}$). For illustration purposes, we first fit the original data using the four different multinomial logit models with main effects, each with proportional odds (*po*) or nonproportional odds (*npo*) assumptions (Bu, Majumdar and Yang, 2020). According to the Akaike information criterion (AIC) (Akaike, 1973), we choose the cumulative logit model with npo:

$$\log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{i5}}\right) = \beta_{j1} + \beta_{j2}x_{i1} + \beta_{j3}x_{i2},$$

with $i = 1, \ldots, 8$ and $j = 1, 2, 3, 4$.

The fitted parameters $\hat{\boldsymbol{\theta}} = (\hat{\beta}_{11}, \hat{\beta}_{21}, \hat{\beta}_{31}, \hat{\beta}_{41}, \hat{\beta}_{12}, \hat{\beta}_{22}, \hat{\beta}_{32}, \hat{\beta}_{42}, \hat{\beta}_{13}, \hat{\beta}_{23}, \hat{\beta}_{33}, \hat{\beta}_{43})^T = (-4.047, -2.225, -0.302, 1.386, 4.214, 3.519, 2.420, 1.284, -0.131, -0.376, -0.237, -0.120)^T$ is used for finding the locally D-optimal allocation $\mathbf{w}_{\mathrm{o}}$ and $\mathbf{n}_{\mathrm{o}}$ for choosing $n = 600$ participants. Since we do not have true parameter values for real data, in order to design EW D-optimal sampler, following Example 5.2 in Bu, Majumdar and Yang (2020), we extract $B = 1000$ bootstrapped samples from the original data and fit the cumulative npo model with bootstrapped samples to obtain randomized parameter vectors $\hat{\boldsymbol{\theta}}_{(1)}, \ldots, \hat{\boldsymbol{\theta}}_{(B)}$ serving as an empirical distribution of $\boldsymbol{\theta}$. Among the fitted parameters by SAS PROC LOGISTIC command, 956 parameter vectors are feasible, that is, in the parameter space $\boldsymbol{\Theta} = \{\boldsymbol{\theta} \in \mathbb{R}^{12} \mid \beta_{j1} + \beta_{j2}x_{i1} + \beta_{j3}x_{i2} < \beta_{j+1,1} + \beta_{j+1,2}x_{i1} + \beta_{j+1,3}x_{i2}, j =$

Table 3. Allocations (Proportions) for Stratified Samplers in Example 8.

Severity	Mild				Severe			
Dose	1	2	3	4	1	2	3	4
Proportional	78 (0.130)	70 (0.117)	75 (0.125)	72 (0.120)	79 (0.132)	72 (0.120)	80 (0.133)	74 (0.123)
Uniform	75 (0.125)	75 (0.125)	75 (0.125)	75 (0.125)	75 (0.125)	75 (0.125)	75 (0.125)	75 (0.125)
Locally D-opt ($\hat{\boldsymbol{\theta}}$)	155 (0.258)	0 (0)	0 (0)	100 (0.167)	168 (0.280)	0 (0)	0 (0)	177 (0.295)
EW D-opt	147 (0.245)	0 (0)	0 (0)	109 (0.182)	168 (0.280)	0 (0)	0 (0)	176 (0.293)

Table 4. Quantiles of Relative Efficiencies in Example 8.

Sampler	Minimum	1st Quartile	Median	3rd Quartile	Maximum
SRSWOR	76.23%	80.16%	80.65%	81.15%	84.11%
Proportional	77.32%	80.33%	80.66%	80.97%	83.39%
Uniform	77.23%	80.05%	80.40%	80.71%	83.13%
EW D-opt	98.91%	99.80%	99.90%	99.96%	100%

$1, 2, 3; i = 1, \ldots, 8\}$ of cumulative logit model (Bu, Majumdar and Yang, 2020, Sec. 5.1). We denote them by $\boldsymbol{\theta}_i$, $i = 1, \ldots, 956$. Then we replace $\mathbf{F}_i$ with $\hat{E}(\mathbf{F}_i) = \sum_{i=1}^{956} \mathbf{F}_i(\boldsymbol{\theta}_i)/956$ to obtain the EW D-optimal allocation $\mathbf{w}_{\text{EW}}$ and $\mathbf{n}_{\text{EW}}$, which maximizes $| \sum_{i=1}^{8} w_i \hat{E}(\mathbf{F}_i) |$. The corresponding allocations are listed in Table 3. In Table 4, we list the quantiles of relative efficiencies of SRSWOR (realized allocations after sampling), proportionally stratified, uniformly stratified, and EW D-optimal allocations with respect to the locally D-optimal allocations based on $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{956}$, respectively. From Table 4, we conclude that in this case, the EW D-optimal sampler is highly efficient compared with the locally D-optimal sampler, and both of them are much more efficient than SRSWOR, proportionally or uniformly stratified sampler.

According to Table 4, the two additional constraints in Example 2, $n(w_1 + w_2 + w_3 + w_4) \leq 392$ and $n(w_5 + w_6 + w_7 + w_8) \leq 410$, are not attained for locally D-optimal and EW D-optimal allocations. In other words, the constrained D-optimal allocations in Example 8 are the same as unconstrained ones in this case. In Example 12 of the Supplementary Material, we provide another example of the sampling problems with the trauma clinical study where the constraints make a difference.

6. Conclusion

In this paper, we consider the constrained subsampling problem for paid research studies or clinical trials to estimate the treatment effects or regression coefficients as accurately as possible. Typically we have some covariates, such

as gender and age, collected along with candidates, which are known to have some influences on the treatment effects. If we do not have any idea about the regression coefficients associated with the covariates, we recommend (constrained) uniformly stratified sampler (Lem. 5); if we have some information about the regression coefficients, such as their signs or ranges, we recommend EW D-optimal sampler; if we have a good idea about the regression coefficients such as estimates from a pilot study, we recommend (locally) D-optimal sampler. We use two examples, one with binary responses and generalized linear models, and the other with 5-category responses and multinomial logistic models, to show that our recommended samplers can be much more efficient than classical samplers for paid research studies or clinical trials. The recommended samplers are fairly robust under model misspecification. We also show that under some circumstances, the constrained uniform sampler is optimal from an experimental design's perspective and can be used as a robust sampling strategy in paid research studies if little information is known about the effects of the covariates.

To find an EW D-optimal sampler or locally D-optimal sampler under constraints, we propose a new algorithm called the constrained lift-one algorithm. While keeping the high efficiency of lift-one algorithms, the proposed algorithm corrects the original lift-one algorithm when additional constraints are added to the design space S. Compared with the constrained optimal designs in the literature, our algorithm can provide highly efficient results for more general statistical models under more general constraints including but not limited to $n_i \leq N_i$.

Supplementary Material

The Supplementary Material includes several sections: S1 describes the lift-one algorithm for general parametric models without constraints; S2 provides a table that lists commonly used GLM models, the corresponding link functions, and ν functions; S3 offers two examples with details in finding r_{i1} and r_{i2}; S4 presents an example assuming different parameter values from Example 6's for robustness with respect to model misspecification; S5 illustrates an example that the D-optimal allocations attain one of the constraints; S6 contains the proofs for lemmas and theorems in this paper.

Acknowledgments

This research is partially supported by the U.S. NSF grant DMS-1924859.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd International Symposium on Information Theory* (Edited by B. Petrov and F. Csaki), 267–281. Akademiai Kiado, Budapest.

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

Atkinson, A., Fedorov, V., Herzberg, A. and Zhang, R. (2014). Elemental information matrices and optimal experimental design for generalized regression models. *Journal of Statistical Planning and Inference* **144**, 81–91.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Bu, X., Majumdar, D. and Yang, J. (2020). D-optimal designs for multinomial logistic models. *The Annals of Statistics* **48**, 983–1000.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science* **10**, 273–304.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics* **24**, 586–602.

Chuang-Stein, C. and Agresti, A. (1997). Tutorial in Biostatistics A review of tests for detecting a monotone dose-response relationship with ordinal response data. *Statistics in Medicine* **16**, 2599–2618.

Cook, D. and Fedorov, V. (1995). Constrained optimization of experimental design. *Statistics* **26**, 129–178.

Dobson, A. and Barnett, A. (2018). *An Introduction to Generalized Linear Models*. 4th Edition. Chapman & Hall/CRC.

Elfving, G. (1952). Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics* **23**, 255–262.

Fang, K.-T. and Wang, Y. (1994). *Number-Theoretic Methods in Statistics*. Chapman and Hall, London.

Fang, K.-T., Wang, Y. and Bentler, P. M. (1994). Some applications of number-theoretic methods in statistics. *Statistical Science* **9**, 416–428.

Fedorov, V. (1972). *Theory of Optimal Experiments*. Academic Press.

Fedorov, V. and Leonov, S. (2014). *Optimal Design for Nonlinear Response Models*. Chapman & Hall/CRC.

Fedorov, V. V. (1989). Optimal design with bounded density: Optimization algorithms of the exchange type. *Journal of Statistical Planning and Inference* **22**, 1–13.

Gao, K., Liu, G. and Tang, W. (2022). High-dimensional reliability analysis based on the improved number-theoretical method. *Applied Mathematical Modelling* **107**, 151–164.

Gong, F., Cui, H., Zhang, L. and Liang, Y. (1999). An improved algorithm of sequential number-theoretic optimization (SNTO) based on clustering technique. *Chemometrics and Intelligent Laboratory Systems* **45**, 339–346.

Jennett, B. and Bond, M. (1975). Assessment of outcome after severe brain damage. *Lancet* **305**, 480–484.

Lay, S. R. (1982). *Convex Sets and Their Applications*. John Wiley & Sons.

Lee, C. M.-S. (1988). Constrained optimal designs. *Journal of Statistical Planning and Inference* **18**, 377–389.

Lohr, S. (2019). *Sampling: Design and Analysis*. Chapman and Hall/CRC.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. 2nd Edition. Chapman and Hall/CRC.

Pronzato, L. (2004). A minimax equivalence theorem for optimum bounded design measures. *Statistics & Probability Letters* **68**, 325–331.

Pronzato, L. (2006). On the sequential construction of optimum bounded designs. *Journal of Statistical Planning and Inference* **136**, 2783–2804.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. John Wiley & Sons.

Silvey, S. (1980). *Optimal Design*. Chapman & Hall/CRC.

Stufken, J. and Yang, M. (2012). Optimal designs for generalized linear models. In *Design and Analysis of Experiments, Volume 3: Special Designs and Applications* (Edited by K. Hinkelmann), 137–165. Wiley.

Welch, W. J. (1982). Branch-and-bound search for experimental designs based on d optimality and other criteria. *Technometrics* **24**, 41–48.

Wynn, H. (1970). The sequential generation of D-optimum experimental designs. *Annals of Mathematical Statistics* **41**, 1655–1664.

Wynn, H. P. (1977a). Minimax purposive survey sampling design. *Journal of the American Statistical Association* **72**, 655–657.

Wynn, H. P. (1977b). Optimum designs for finite populations sampling. In *Statistical Decision Theory and Related Topics*, 471–478. Elsevier.

Wynn, H. P. (1982). Optimum submeasures with application to finite population sampling. In *Statistical Decision Theory and Related Topics III*, 485–495.

Yang, J. and Mandal, A. (2015). D-optimal factorial designs under generalized linear models. *Communications in Statistics - Simulation and Computation* **44**, 2264–2277.

Yang, J., Mandal, A. and Majumdar, D. (2012). Optimal designs for two-level factorial experiments with binary response. *Statistica Sinica* **22**, 885–907.

Yang, J., Mandal, A. and Majumdar, D. (2016). Optimal designs for 2^k factorial experiments with binary response. *Statistica Sinica* **26**, 385–411.

Yang, J., Tong, L. and Mandal, A. (2017). D-optimal designs with ordered categorical data. *Statistica Sinica* **27**, 1879–1902.

Zocchi, S. and Atkinson, A. (1999). Optimum experimental designs for multinomial logistic models. *Biometrics* **55**, 437–444.

(Received December 2022; accepted August 2023)