Statistica Sinica **26** (2016), 385-411
doi:http://dx.doi.org/10.5705/ss.2013.265

OPTIMAL DESIGNS FOR 2^k FACTORIAL EXPERIMENTS WITH BINARY RESPONSE

Jie Yang[1], Abhyuday Mandal[2] and Dibyen Majumdar[1]

[1] *University of Illinois at Chicago and* [2] *University of Georgia*

Abstract: We consider the problem of obtaining D-optimal designs for factorial experiments with a binary response and k qualitative factors each at two levels. We obtain a characterization of locally D-optimal designs. We then develop efficient numerical techniques to search for locally D-optimal designs. Using prior distributions on the parameters, we investigate EW D-optimal designs that maximize the determinant of the expected information matrix. It turns out that these designs can be obtained easily using our algorithm for locally D-optimal designs and are good surrogates for Bayes D-optimal designs. We also investigate the properties of fractional factorial designs and study robustness with respect to the assumed parameter values of locally D-optimal designs.

Key words and phrases: D-optimality, EW D-optimal design, fractional factorial design, full factorial design, generalized linear model, uniform design.

1. Introduction

Our goal is to determine optimal and efficient designs for factorial experiments with qualitative factors and a binary response. The traditional factorial design literature deals with experiments where the factors have discrete levels and the response follows a linear model (see, for example, Xu, Phoa, and Wong (2009) and references therein). On the other hand, there is a growing body of literature on optimal designs for quantitative factors with binary or categorical response. For the specific experiments we study, however, the design literature is meager. Consequently, these experiments are usually designed by the guidelines of traditional factorial design theory for linear models. As we shall see, the resulting designs can be quite inefficient, especially when compared to designs that make use of prior information when it is available. Our goal is to address this problem directly and determine efficient designs specifically for experiments with qualitative factors and a binary response.

We assume that the process under study is adequately described by a generalized linear model (GLM). GLMs have been widely used for modeling binary response. Stufken and Yang (2012) noted that "the study of optimal designs for experiments that plan to use a GLM is however not nearly as well developed

(see also Khuri, Mukherjee, Sinha, and Ghosh (2006)), and tends to be much more difficult than the corresponding and better studied problem for the special case of linear models." For optimal designs under GLMs, there are four different approaches proposed in the literature to handle the dependence of the design optimality criterion on the unknown parameters: the local optimality approach of Chernoff (1953) in which the parameters are replaced by assumed values; the Bayesian approach (Chaloner and Verdinelli (1995)) that incorporates prior belief on unknown parameters; the maximin approach that maximizes the minimum efficiency over a range of values of the unknown parameters (see Pronzato and Walter (1988) and Imhof (2001)); the sequential approach where the design and parameter estimates are updated in an iterative way (see Ford, Titterington, and Kitsos (1989)). We focus on local optimality and study D-optimal factorial designs under GLMs. We also consider Bayes optimality and study a surrogate for Bayes D-optimal designs that has many desirable properties.

The methods for analyzing data from GLMs have been discussed in depth in the literature (for example, McCullagh and Nelder (1989), Agresti (2013), Lindsey (1997), McCulloch and Searle (2001), Dobson and Barnett (2008), and Myers, Montgomery, and Vining (2002)). Khuri, Mukherjee, Sinha, and Ghosh (2006) provided a systematic study of the optimal design problem in the GLM setup and recently there has been an upsurge in research in both theory and computation of optimal designs. Russell et al. (2009), Li and Majumdar (2008, 2009), Yang and Stufken (2009), Yang, Zhang, and Huang (2011), Stufken and Yang (2012) are some of the papers that developed theory, while Woods et al. (2006), Dror and Steinberg (2006, 2008), Waterhouse et al. (2008), Woods and van de Ven (2011) focused on developing efficient numerical techniques for obtaining optimal designs under generalized linear models. Our focus is on optimal designs for GLMs with qualitative factors.

The case of 2^2 experiments with qualitative factors and a binary response was studied by Yang, Mandal, and Majumdar (2012), where we obtained optimal designs analytically in special cases, and demonstrated how to obtain a solution in the general case using cylindrical algebraic decomposition. The optimal allocations were shown to be robust to the choice of the assumed values of the model parameters. Graßhoff and Schwabe (2008) has some relevant results for the $k = 2$ factor case. The extension for $k > 2$ factors is substantial due to additional complexities associated with determination, computation, and robustness of optimal designs that are not present in the two-factor case. For the general case of 2^k experiments with binary response, Dorta-Guerra, González-Dávila, and Ginebra (2008) obtained an expression for the D-criterion and studied several special cases.

A motivating example is the odor removal study conducted by textile engineers at the University of Georgia. The scientists studied the manufacture of

Table 1. Factors and levels, odor experiment.

Factor	Levels	−	+
A	algae	raffinated or solvent extracted algae	catfish pond algae
B	scavenger material	Aqua Tech activated carbon	BYK-P 4200 purchased from BYK Additives Instruments
C	synthetic resin	polyethylene	polypropylene
D	compatabilizers	absent	present

bio-plastics from algae that contain odorous volatiles. These odorous volatiles, generated from algae bio-plastics, either occur naturally within the algae or are generated through the thermoplastic processing due to heat and pressure. In order to commercialize these algae bio-plastics, the odor causing volatiles must be removed. Static headspace microextraction and gas chromatography − mass spectroscopy are used to identify the odorous compounds and qualitatively assess whether or not they have been successfully removed. The outcome of this assessment is the response of the experiment. For that purpose, a study was conducted with a 2^{4-1}_{IV} design, a regular fraction, with five replicates using algae and synthetic plastic resin blends. The four different factors were: type of algae, scavenger material (adsorbent), synthetic resin and compatabilizers (see Table 1 for details).

We obtain theoretical results and algorithms for locally optimal designs for k qualitative factors at two levels each and a binary response in the generalized linear model setup. We consider D-optimal designs maximizing the determinant of the information matrix. Although we explore designs in which observations are taken at every possible level combination, when the number of factors is large, such full factorials are practically infeasible. Hence the study of fractional factorial designs occupies a substantial part of the linear-model based design literature, and we too study these designs in our setup. A natural question that arises when we use local optimality is whether the resulting designs are robust to the assumed parameter values. We consider this in Section 5.

An alternative approach to design optimality is Bayes optimality (Chaloner and Verdinelli (1995)). For our problem, however, for large k ($k \geq 4$) the computations quickly become expensive. Hence, as a surrogate criterion, we explore a D-optimality criterion with the information matrix replaced by its expectation under the prior. This is one of the suggested alternatives to formal Bayes optimality in Atkinson, Donev, and Tobias (2007). It has been used by Zayats and Steinberg (2010) for optimal designs for detection capability of networks. We call this *EW D-optimality* (E for expectation, W for the notation w_i used for the GLM "weight", which can be thought of as information contained in an

individual observation). Effectively this reduces to a locally optimal design with local values of the weight parameters replaced by their expectations. The EW D-optimal designs are very good and easy-to-compute surrogates for Bayes D-optimal designs. Unless k is small or the experimenter is quite certain about the parameter values, we suggest the use of EW D-optimal designs. The use of surrogates of Bayes optimality has been recommended by Gotwalt, Jones, and Steinberg (2009).

Beyond theoretical results, the question that may be asked is whether these results give the user any advantage in concrete experiments. It turns out that when $k > 2$, in most situations, we gain considerably by taking advantage of our results instead of using standard linear-model results. Unlike the linear model case, not all nonsingular regular fractions have the same D-efficiency. Indeed, if we have some knowledge of the parameters, we are able to identify an efficient fractional factorial design, that is often not a regular fraction.

This paper is organized as follows. In Section 2 we describe the preliminary setup. In Section 3 we provide several results for locally D-optimal designs, including their uniqueness, a characterization of locally D-optimal design, the concept of EW D-optimal designs, and algorithms for finding D-optimal designs. In Section 4 we discuss the properties of fractional factorial designs. We address the robustness of D-optimal designs in Section 5, and revisit the odor example in Section 6. Some concluding remarks and topics for future research are discussed in Section 7. Additional results, proofs and some details on the algorithms are relegated to the Supplementary Materials.

2. Preliminary Setup

Consider an experiment with a binary response and k explanatory variables at 2 levels each. Suppose n_i units are allocated to the ith experimental condition such that $n_i \geqslant 0$, $i = 1, \ldots, 2^k$, and $n_1 + \cdots + n_{2^k} = n$. We suppose that n is fixed and the problem is to determine the "optimal" n_i's. In fact, we write our optimality criterion in terms of the proportions

$$p_i = \frac{n_i}{n}, \quad i = 1, \ldots, 2^k$$

and determine the "optimal" $p_i \geq 0$ satisfying $\sum_{i=1}^{2^k} p_i = 1$. Since n_i's are integers, an optimal design obtained in this fashion may not always be viable. In Section 3.3.2 we consider the design problem over integer n_i's.

We use a generalized linear model setup. Suppose η is a linear predictor that involves the main effects and the interactions that are assumed to be in the model. For instance, for a 2^3 experiment with a model that includes the main effects and the two-factor interaction of factors 1 and 2, $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2$, where each $x_i \in \{-1, 1\}$. The aim of the experiment is to obtain inferences about

the parameter vector of factor effects $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_{12})'$. In the framework of generalized linear models, the expectation of the response Y, $E(Y) = \pi$, is connected to the linear predictor η by the link function g: $\eta = g(\pi)$ (McCullagh and Nelder (1989)). For a binary response, the commonly used link functions are logit, probit, log-log, and complementary log-log links.

The maximum likelihood estimator of $\boldsymbol{\beta}$ has an asymptotic covariance matrix (McCullagh and Nelder (1989); Khuri, Mukherjee, Sinha, and Ghosh (2006)) that is the inverse of $nX'WX$, where $W = \text{diag}\{w_1 p_1, \ldots, w_{2^k} p_{2^k}\}$, $w_i = (d\pi_i/d\eta_i)^2 / (\pi_i(1-\pi_i)) \geq 0$, η_i and π_i correspond to the ith experimental condition for η and π, and X is the "model matrix". For example, for a 2^3 experiment with model $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2$,

$$X = \begin{pmatrix} +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & -1 & +1 \end{pmatrix}. \tag{2.1}$$

The n_i's determine how many observations are made at each experimental condition, which are characterized by the rows of X. A D-optimal design maximizing $|X'WX|$ depends on the w_i's, which in turn depend on the regression parameters $\boldsymbol{\beta}$ and the link function g. We discuss D-optimal designs in terms of w_i's so that our results are not limited to specific link functions.

Unlike experiments with continuous factors, the 2^k design points in our setup are fixed and we only have the option of determining the optimal proportions. For results on optimal designs with continuous factors in the GLM setup see, for example, Stufken and Yang (2012).

3. Locally D-Optimal Designs

We start with a formulation of the local D-optimality problem and establish some general results. Consider a 2^k experiment. The goal is to find an optimal $\mathbf{p} = (p_1, p_2, \ldots, p_{2^k})'$ which maximizes $f(\mathbf{p}) := |X'WX|$ for specified values of $w_i \geq 0, i = 1, \ldots, 2^k$. The specification of the w_i's come from the initial values of the parameters and the link function. Here $p_i \geq 0$, $i = 1, \ldots, 2^k$, and $\sum_{i=1}^{2^k} p_i = 1$. It is easy to see that there always exists a D-optimal allocation $\mathbf{p}$ since the set of all feasible allocations is bounded and closed. On the other hand, the uniqueness of D-optimal designs is usually not guaranteed (see Remark 2). Even if all the p_i's are positive, the resulting design is not full factorial in the traditional sense

where equal numbers of replicates are used. If some of the p_i's are zero, it is a fractional factorial design; these will be discussed in the next section. In fact, the number of nonzero p_i's in the optimal design could be much less than 2^k, as we will see in Section 3.3.

3.1. Characterization of locally D-optimal designs

Suppose the parameters (main effects and interactions) are $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)'$, where $d \geq k$. The following lemma expresses the objective function as an order-$(d+1)$ homogeneous polynomial of $p_1, \ldots, p_{2^k}$.

Lemma 1. *Let $X[i_1, i_2, \ldots, i_{d+1}]$ be the $(d+1) \times (d+1)$ sub-matrix consisting of the i_1th, i_2th, ..., i_{d+1}th rows of the model matrix X. Then*

$$f(\mathbf{p}) = |X'WX| = \sum_{1 \leq i_1 < \cdots < i_{d+1} \leq 2^k} |X[i_1, i_2, \ldots, i_{d+1}]|^2 \cdot p_{i_1} w_{i_1} p_{i_2} w_{i_2} \cdots p_{i_{d+1}} w_{i_{d+1}}.$$

González-Dávila, Dorta-Guerra, and Ginebra (2007, Proposition 2.1) obtained essentially the same result. This can also be proved directly using the results of Rao (1973, Chap. 1). From Lemma 1 it is immediate that at least $(d+1)$ w_i's, as well as the corresponding p_i's, have to be positive for the determinant $f(\mathbf{p})$ to be nonzero. This implies that if $\mathbf{p}$ is D-optimal, then $p_i < 1$ for each i. Theorem 1 below gives a sharper bound, $p_i \leq 1/(d+1)$ for each $i = 1, \ldots, 2^k$, for the optimal allocation. Define, for each $i = 1, \ldots, 2^k$,

$$f_i(z) = f\left(\frac{1-z}{1-p_i}p_1, \ldots, \frac{1-z}{1-p_i}p_{i-1}, z, \frac{1-z}{1-p_i}p_{i+1}, \ldots, \frac{1-z}{1-p_i}p_{2^k}\right), \quad 0 \leq z \leq 1. \tag{3.1}$$

Then $f_i(z)$ is well defined for all $\mathbf{p}$ of interest.

Theorem 1. *If $f(\mathbf{p}) > 0$, $\mathbf{p}$ is D-optimal if and only if for each $i = 1, \ldots, 2^k$, one of the following is satisfied:*

(i) $p_i = 0$ *and* $f_i(1/2) \leq [(d+2)/(2^{d+1})] f(\mathbf{p})$;

(ii) $0 < p_i \leq 1/(d+1)$ *and* $f_i(0) = [(1 - p_i(d+1))/((1-p_i)^{d+1})] f(\mathbf{p})$.

Remark 1. Theorem 1 is essentially a specialized version of the general equivalence theorem on a pre-determined finite set of design points. Unlike the usual form of the equivalence conditions (for examples, see Kiefer (1974), Pukelsheim (1993), Atkinson, Donev, and Tobias (2007), Stufken and Yang (2012), Fedorov and Leonov (2014)) where the inverse matrix of $X'WX$ needs to be calculated, Theorem 1 is expressed in terms of the quantities $f(\mathbf{p})$, $f_i(1/2)$ and $f_i(0)$ only. These expressions are critical for the algorithms proposed later. The theorem also gives a sharper bound, $0 < p_i \leq 1/(d+1)$, for support points. Note that even if

$p_i = 0$ for some i, it is still possible that the equality $f_i(1/2) = (d+2)/(2^{d+1}) \cdot f(\mathbf{p})$ holds. In the Supplementary Materials, we provide a self-contained proof of Theorem 1 that does not rely on any general equivalence theorem. Its connection to the general equivalence theorem is provided in the Supplementary Materials.

Designs that are supported on $(d+1)$ points are attractive in many experiments because they require a minimum number of settings. In our context, a design $\mathbf{p} = (p_1, \ldots, p_{2^k})'$ is called *minimally supported* if it has exactly $(d+1)$ nonzero p_i's. For designs supported on rows $i_1, \ldots, i_{d+1}$, the D-optimal choice of weights is $p_{i_1} = \cdots = p_{i_{d+1}} = 1/(d+1)$. This result can be obtained from Lemma 1 directly. Yang, Mandal, and Majumdar (2012) found a necessary and sufficient condition for a minimally supported design to be D-optimal for 2^2 main-effects model. With the aid of Theorem 1, we provide a generalization for 2^k designs in the next theorem. Here $w_i > 0$ for each i for the commonly used link functions, including logit, probit, and (complementary) log-log.

Theorem 2. *Assume $w_i > 0$, $i = 1, \ldots, 2^k$. Let $\mathbf{I} = \{i_1, \ldots, i_{d+1}\} \subset \{1, \ldots, 2^k\}$ be an index set satisfying $|X[i_1, \ldots, i_{d+1}]| \neq 0$. Then the minimally supported design satisfying $p_{i_1} = p_{i_2} = \cdots = p_{i_{d+1}} = 1/(d+1)$ is D-optimal if and only if for each $i \notin \mathbf{I}$,*

$$\sum_{j \in \mathbf{I}} \frac{|X[\{i\} \cup \mathbf{I} \setminus \{j\}]|^2}{w_j} \leq \frac{|X[i_1, i_2, \ldots, i_{d+1}]|^2}{w_i}.$$

For example, under the 2^2 main-effects model, since $|X[i_1, i_2, i_3]|^2$ is constant across all choices of i_1, i_2, i_3, $p_1 = p_2 = p_3 = 1/3$ is D-optimal if and only if $v_1 + v_2 + v_3 \leq v_4$, where $v_i = 1/w_i$, $i = 1, 2, 3, 4$. This gives us Theorem 1 of Yang, Mandal, and Majumdar (2012). For the 2^3 main-effects model, the model matrix X is given by (2.1) with the last column deleted. Using this order of rows, the standard regular fractional factorial design $p_1 = p_4 = p_6 = p_7 = 1/4$ given by the defining relation $1 = ABC$ is D-optimal if and only if $v_1 + v_4 + v_6 + v_7 \leq 4 \min\{v_2, v_3, v_5, v_8\}$, and the other standard regular fractional design $p_2 = p_3 = p_5 = p_8 = 1/4$ is D-optimal if and only if $v_2 + v_3 + v_5 + v_8 \leq 4 \min\{v_1, v_4, v_6, v_7\}$.

Remark 2. In order to characterize the uniqueness of the optimal allocation, we define a matrix $X_w = [\mathbf{1},\ \mathbf{w} * \mathbf{1},\ \mathbf{w} * \gamma_2,\ \ldots,\ \mathbf{w} * \gamma_s]$, where $\mathbf{1}$ is the $2^k \times 1$ vector of all 1's, $\{\mathbf{1}, \gamma_2, \ldots, \gamma_s\}$ forms the set of all distinct pairwise Schur products (or entrywise product) of the columns of the model matrix X, $\mathbf{w} = (w_1, \ldots, w_{2^k})'$, and "$*$" indicates Schur product. It can be verified that any two feasible allocations ($p_i \geq 0$ satisfying $\sum_{i=1}^{2^k} p_i = 1$) generate the same matrix $X'WX$ as long as the difference of the matrices belongs to the null space of X_w. If $\text{rank}(X_w) < 2^k$, any criterion based on $X'WX$ yields an affine set of solutions

with dimension $2^k - \text{rank}(X_w)$. If $\text{rank}(X_w) = 2^k$, the D-optimal allocation $\mathbf{p}$ is unique. For example, for a 2^3 design the model consisting of all main effects and one two-factor interaction, or for a 2^4 design the model consisting of all main effects, all two-factor interactions, and one three-factor interaction, the D-optimal allocation is unique.

3.2. EW D-optimal designs

Since locally D-optimal designs depend on w_i's, they require assumed values of w_i's, or β_i's, as input. In Section 5, we examine the robustness of D-optimal designs to mis-specification of β_i's. An alternate to local optimality is Bayes optimality (Chaloner and Verdinelli (1995)). In our setup, a Bayes D-optimal design maximizes $E(\log|X'WX|)$ where the expectation is taken over the prior on β_i's. One difficulty of Bayes optimality is that it is computationally expensive. To overcome this drawback we explore an alternative suggested by Atkinson, Donev, and Tobias (2007) where W in the Bayes criterion is replaced by its expectation. We call this *EW* D-optimality where *EW* stands for expectation of W.

Definition. An *EW D-optimal* design is an optimal allocation $\mathbf{p}$ that maximizes $|X'E(W)X|$.

Note that EW D-optimality may be viewed as local D-optimality with the w_i's replaced by their expectations. All existence and uniqueness properties of locally D-optimal design apply. Since $w_i > 0$ for all $\boldsymbol{\beta}$ under typical link functions, $E(w_i) > 0$ for each i. By Jensen's inequality,

$$E\left(\log|X'WX|\right) \leq \log|X'E(W)X|$$

since $\log|X'WX|$ is concave in $\mathbf{w}$. Thus an EW D-optimal design maximizes an upper bound for Bayesian D-optimality criterion.

In practice, once the $E(w_i)$'s are calculated via numerical integration, algorithms for local D-optimality can be applied with w_i replaced by $E(w_i)$. We will show that EW D-optimal designs are often almost as efficient as designs that are optimal with respect to the Bayes D-optimality criterion, while realizing considerable savings in computation time. In fact, while searching for an EW D-optimal design, integration can be performed in advance of the optimization. This provides a computational advantage over the search for Bayesian D-optimal designs, where integration needs to be performed in each step of the optimization in order to evaluate the design. Furthermore, EW D-optimal designs are highly robust in terms of maximum loss of efficiency (Section 5).

Given the link function g, let $\nu = \left[\left(g^{-1}\right)'\right]^2 / \left[g^{-1}(1-g^{-1})\right]$. Then $w_i = \nu(\eta_i) = \nu\left(\mathbf{x}_i'\boldsymbol{\beta}\right)$, $i = 1, \ldots, 2^k$, where $\mathbf{x}_i$ is the ith row of the model matrix X, and

$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)'$. If the regression coefficients $\beta_0, \beta_1, \ldots, \beta_d$ are independent, and $\beta_1, \ldots, \beta_d$ each has a symmetric distribution about 0 (not necessarily the same distribution) then all the $w_i,\ i = 1, \ldots, 2^k$ have the same distribution and the uniform design $p_1 = \cdots = p_{2^k} = 2^{-k}$ is an EW D-optimal design for any given link function (by "uniform design" we mean a design with uniform allocation on its support points). On the other hand, in many experiments we may be able to assume that the slope of a main effect is non-decreasing. If $\beta_i \in [0, \beta_{iu}]$ for each i, the uniform design will not be EW D-optimal in general, as illustrated in the following example.

Example 1. Consider a 2^3 experiment with main-effects model. Suppose β_0, β_1, β_2 and β_3 are independent, $\beta_0 \sim U[-3, 3]$, and $\beta_1, \beta_2, \beta_3 \sim U[0, 3]$. Then $E(w_1) = E(w_8) = 0.042$, $E(w_2) = E(w_3) = \cdots = E(w_7) = 0.119$. Under the logit link the EW D-optimal design is $\mathbf{p}_e = (0, 1/6, 1/6, 1/6,\ 1/6,\ 1/6,\ 1/6,\ 0)'$, and the Bayesian D-optimal design, which maximizes $\phi(\mathbf{p}) = E(\log|X'WX|)$, is $\mathbf{p}_o = (0.004,\ 0.165,\ 0.166,\ 0.165,\ 0.165,\ 0.166,\ 0.165,\ 0.004)'$. The efficiency of $\mathbf{p}_e$ with respect to $\mathbf{p}_o$ is $\exp\{(\phi(\mathbf{p}_e) - \phi(\mathbf{p}_o))/(d+1)\} \times 100\% = 99.98\%$, while the efficiency of the uniform design is 94.39%. Here the EW and Bayes criteria lead to virtually the same design. It is remarkable that it takes 2.39 seconds to find an EW solution while it takes 121.73 seconds to find a Bayes solution. The difference in computational time is even more prominent for the 2^4 case (24 seconds versus 3,147 seconds). All multiple integrals here are calculated using `R` function `adaptIntegrate` in the package `cubature`.

3.3. Algorithms to search for locally D-optimal allocation

In this section, we develop efficient algorithms to search for locally D-optimal allocations with given w_i's. The same algorithms can be used for finding EW D-optimal designs.

3.3.1. Lift-one algorithm for maximizing $f(\mathbf{p}) = |X'WX|$

We propose the *lift-one* algorithm for obtaining locally D-optimal $\mathbf{p} = (p_1, \ldots, p_{2^k})'$ with given w_i's. The basic idea is that, for randomly chosen $i \in \{1, \ldots, 2^k\}$, we update p_i to p_i^* and all the other p_j's to $p_j^* = p_j \cdot [(1 - p_i^*)/(1 - p_i)]$. This technique is motivated by the coordinate descent algorithm (Zangwill (1969)). It is also in spirit similar to one-point correction in the literature (Wynn (1970); Fedorov (1972); Müller (2007)), where design points are added/adjusted one by one. The major advantage of the lift-one algorithm is that in order to determine an optimal p_i^*, we need to calculate $|X'WX|$ only once due to Lemma 1 (see Step 3° of the algorithm below).

The lift-one algorithm:

1° Start with arbitrary $\mathbf{p}_0 = (p_1, \ldots, p_{2^k})'$ satisfying $0 < p_i < 1$, $i = 1, \ldots, 2^k$ and compute $f(\mathbf{p}_0)$.

2° Set up a random order for i, going through $\{1, 2, \ldots, 2^k\}$.

3° Following the random order of i in 2°, for each i, determine $f_i(z)$ as in (S.2) in the Supplementary Materials. In this step, either $f_i(0)$ or $f_i(1/2)$ needs to be calculated according to (3.1).

4° Take $\mathbf{p}_*^{(i)} = ([(1 - z_*)/(1 - p_i)]p_1, \ldots, [(1 - z_*)/(1 - p_i)]p_{i-1}, z_*, [(1 - z_*)/(1 - p_i)]p_{i+1}, \ldots, [(1 - z_*)/(1 - p_i)]p_{2^k})'$, where z_* maximizes $f_i(z)$ with $0 \le z \le 1$ (see Lemma S1.3). Then $f(\mathbf{p}_*^{(i)}) = f_i(z_*)$. Lemma S1.3 provides the update in terms of $f_i(0)$ or $f_i(1/2)$.

5° Replace $\mathbf{p}_0$ with $\mathbf{p}_*^{(i)}$, $f(\mathbf{p}_0)$ with $f(\mathbf{p}_*^{(i)})$.

6° Repeat 2° $\sim$ 5° until convergence, $f(\mathbf{p}_0) = f(\mathbf{p}_*^{(i)})$ for each i.

While in all examples that we studied, the lift-one algorithm converges quickly, we do not have a proof of convergence. There is a modified lift-one algorithm, which is only slightly slower, that can be shown to converge, described as follows. For the $10m$th iteration and a fixed order of $i = 1, \ldots, 2^k$ we repeat steps 3° $\sim$ 5°, $m = 1, 2, \ldots$. If $\mathbf{p}_*^{(i)}$ is a better allocation found by the lift-one algorithm than $\mathbf{p}_0$, instead of updating $\mathbf{p}_0$ to $\mathbf{p}_*^{(i)}$ immediately, we obtain $\mathbf{p}_*^{(i)}$ for each i, and replace $\mathbf{p}_0$ with the first best one among $\left\{\mathbf{p}_*^{(i)}, i = 1, \ldots, 2^k\right\}$. The updating strategy at the $10m$th iteration here is similar to the Fedorov-Wynn algorithm (Fedorov (1972), Fedorov and Hackl (1997)) but with a more efficient updating formula. For iterations other than the $10m$th, we follow the original lift-one algorithm update.

Theorem 3. *When the lift-one algorithm or the modified lift-one algorithm converges, the resulting allocation* $\mathbf{p}$ *maximizes* $|X'WX|$ *on the set of feasible allocations. Furthermore, the modified lift-one algorithm is guaranteed to converge.*

Our simulation studies indicate that as k grows, the optimal designs produced by the lift-one algorithm for main-effects models is supported only on a fraction of all the 2^k design points. To illustrate this, we randomly generated the regression coefficients i.i.d. from $U(-3, 3)$ and applied our algorithm to find the optimal designs under the logit link. Figure 1 gives histograms of the numbers of support points in optimal designs found by the lift-one algorithm. Thus, with $k = 2$, 76% of the designs are supported on three points and 24% of them are supported on all

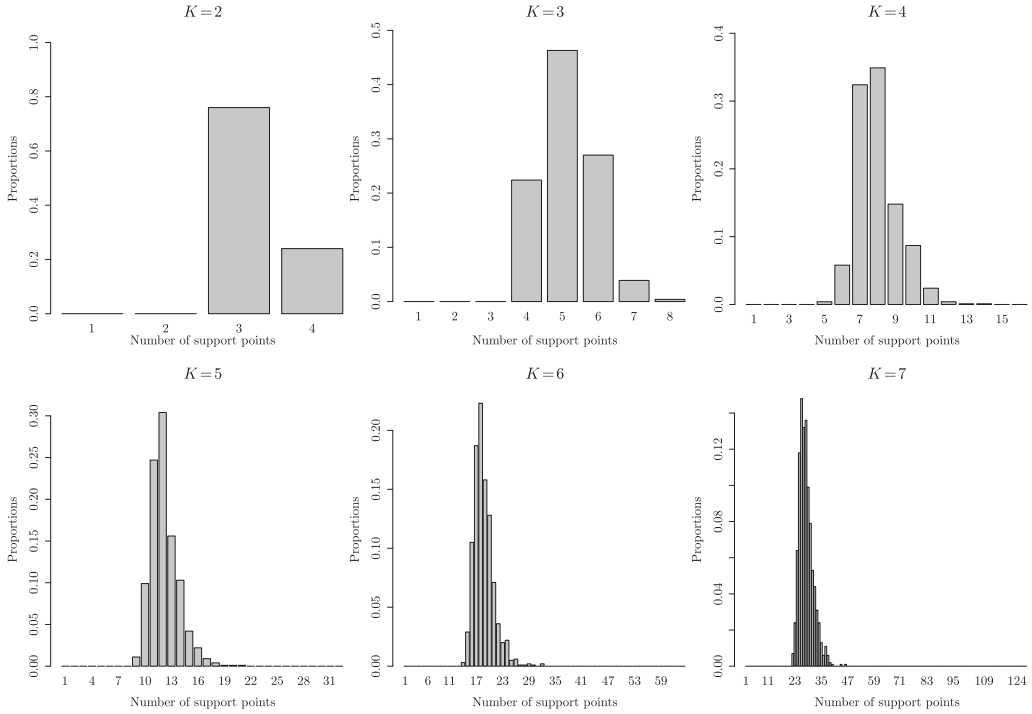

Figure 1. Number of support points in an optimal design (based on 1,000 simulations).

four points. For larger k, the number of support points moves toward a smaller fraction of 2^k. On the other hand, a narrower range of coefficients requires a larger portion of support points. For example, the mean numbers of support points with β_i's i.i.d. from $U(-3,3)$ are $3.2, 5.1, 8.0, 12.4, 18.7, 28.2$ for $k = 2, 3, 4, 5, 6, 7$, respectively. The corresponding numbers increase to $4.0, 7.1, 11.9, 19.1, 30.6, 47.7$ for $U(-1,1)$, and further to $4.0, 7.6, 14.1, 24.7, 41.2, 66.8$ for $U(-0.5, 0.5)$.

The lift-one algorithm is much faster than commonly used optimization techniques (Table 2), including Nelder-Mead, quasi-Newton, conjugate-gradient, simulated annealing (for a comprehensive reference, see Nocedal and Wright (1999)), as well as popular design algorithms for similar purposes including the Fedorov-Wynn (Fedorov (1972), Fedorov and Hackl (1997), Fedorov and Leonov (2014)), Multiplicative (Titterington (1976, 1978), Silvey, Titterington, and Torsney (1978)), and Cocktail (Yu (2010)) algorithms. We utilized the function `constrOptim` in `R` to implement Nelder-Mead, quasi-Newton, conjugate-gradient, and simulated annealing algorithms. As the number of design points (2^k) increases, those algorithms fail to achieve adequately accurate solutions (marked by "−" in Table 2, indicating that the relative efficiency compared

Table 2. Performance of the lift-one algorithm (CPU time in seconds for 100 simulated $\boldsymbol{\beta}$ from U$(-3, 3)$ with logit link and main-effects model).

	Algorithms							
Designs	Nelder-Mead	quasi-Newton	conjugate gradient	simulated annealing	Fedorov -Wynn	Multiplicative	Cocktail	Proposed lift-one
2^2	1.42	0.19	2.09	83.09	6.14	0.28	0.16	0.11
2^3	8.76	24.64	171.74	18.54	11.25	0.86	0.53	0.36
2^4	17.88	–	–	–	21.77	10.97	4.46	1.07
2^5	31.64	–	–	–	47.66	50.12	68.88	4.82
2^6	–	–	–	–	106.89	229.17	189.83	18.29
2^7	–	–	–	–	241.80	890.44	439.55	75.58

with the lift-one solutions is below 80% on average). Thus it takes the Nelder-Mead algorithm 51.73 seconds to find solutions ($\mathbf{p}_{NM}$) at $k = 6$ whose relative efficiency, $(f(\mathbf{p}_{NM})/f(\mathbf{p}_{lo}))^{1/(k+1)}$, compared with the lift-one solutions ($\mathbf{p}_{lo}$) is only 65% on average. As k increases from 2 to 3, although the time spent for simulated annealing algorithm reduces from 83.09 seconds to 18.54 seconds, the relative efficiency on average decreases from 99.8% to 93.0% (it drops down to 66% at k=4 and 54% at k=5). The relative efficiencies do not improve much if more iterations or multiple initial points are allowed. The implementation of the Fedorov-Wynn algorithm here is mainly based on Fedorov and Leonov (2014, §3.1) with updating formula for $(X'WX)^{-1}$. As for the Multiplicative and Cocktail algorithms, we followed Yu (2010) and Mandal, Wong, and Yu (2015). Each of these algorithms achieves essentially the same efficiency compared to the lift-one algorithm. For a fair comparison, the programs were written in `R`, controlled by the same relative convergence tolerance 10^{-5}, and run at the same computer with Intel CPU at 2.5GHz, 8GB memory, and 64-bit (Windows 8.1) Operating System. Based on the simulation results shown in Table 2, the lift-one algorithm runs at a much faster speed across different model setups. In terms of the number of support points on average, only the solutions found by the Cocktail algorithm are comparable with lift-one solutions. Typically, the Multiplicative algorithm finds twice as many support points as does the lift-one, while the other five algorithms simply keep positive weights on all 2^k design points.

Remark 3. There are at least two advantages of the proposed algorithm over its competitors: it exploits the convex structure of the optimization problem (the set of design measures over $\{-1, 1\}^k$ is convex, and the objective function $f(\mathbf{p})$ is log-concave), whereas some of the other algorithms compared do not; it reduces the number of determinant calculations required per iteration of the algorithm. In Table 2 the comparison with a Federov-Wynn algorithm demonstrates that the gain in speed due to these features of the new algorithm is significant.

3.3.2. Algorithm for maximizing $|X'WX|$ with integer solutions

To maximize $|X'WX|$, we propose an *exchange algorithm* that adjusts p_i and p_j simultaneously for a randomly chosen index pair (i, j) (see the Supplementary Materials for detailed description). The original idea of exchange was suggested by Fedorov (1972). It follows from Lemma 1 that the optimal adjusted (p_i^*, p_j^*) can be obtained easily by maximizing a quadratic function. Unlike the lift-one algorithm, the exchange algorithm can be applied to search for integer-valued optimal allocation $\mathbf{n} = (n_1, \ldots, n_{2^k})'$, where $\sum_i n_i = n$.

The exchange algorithm:

1° Start with initial design $\mathbf{n} = (n_1, \ldots, n_{2^k})'$ such that $f(\mathbf{n}) > 0$.

2° Set up a random order of (i, j) going through all pairs.

3° For each (i, j), let $m = n_i + n_j$. If $m = 0$, let $\mathbf{n}_{ij}^* = \mathbf{n}$. Otherwise, calculate $f_{ij}(z)$ at (S.5). Let

$$\mathbf{n}_{ij}^* = (n_1, \ldots, n_{i-1}, z_*, n_{i+1}, \ldots, n_{j-1}, m - z_*, n_{j+1}, \ldots, n_{2^k}),$$

where the integer z_* maximizes $f_{ij}(z)$ with $0 \le z \le m$ according to Lemma S1.5 in the Supplementary Materials. Now $f(\mathbf{n}_{ij}^*) = f_{ij}(z_*) \ge f(\mathbf{n}) > 0$.

4° Repeat 2° $\sim$ 3° until convergence (no more increase in terms of $f(\mathbf{n})$ by any pairwise adjustment).

As expected, the integer-valued optimal allocation $(n_1, \ldots, n_{2^k})'$ is consistent with the proportion-valued allocation $(p_1, \ldots, p_{2^k})'$ for large n. For small n, the algorithm may be used for the fractional design problem in Section 4. The exchange algorithm for integer-valued solutions is not guaranteed to converge to the optimal solutions, especially when n is small compared to 2^k. However, when we search for optimal proportions our algorithm, with slight modification, is guaranteed to converge (see the Supplementary Materials for details).

In terms of finding optimal proportions, the exchange algorithm produces essentially the same results as the lift-one algorithm, although the former is slower. For example, based on 1,000 simulated $\boldsymbol{\beta}$'s from U(-3,3) with logit link and the main-effects model, the ratio of computational time of the exchange algorithm over the lift-one algorithm is 6.2, 10.2, 16.8, 28.8, 39.5 and 51.3 for $k = 2, \ldots, 7$, respectively. It requires 2.02, 5.38, 19.2, 84.3, 352, and 1,245 seconds, respectively, to finish the 1,000 simulations using the lift-one algorithm on a regular PC with 2.26GHz CPU and 2.0G memory.

The general purpose optimization algorithms might be a little slow and faster alternatives should exist. Thus, the adaptive barrier method might be inefficient compared to transformations to obtain an unconstrained optimization problem.

For the pseudo-Bayesian designs, it is possible that a fixed quadrature scheme would be faster, though possibly less accurate. Detailed study of the computational properties of the proposed algorithms is a topic for future research.

4. Fractional Factorial Designs

If for the optimal allocation some p_i's are zero, then the resulting design is necessarily a fractional factorial. Even if all of the proportions in the optimal design are substantially away from zero, the experimenter may need, or prefer, to use a fractional factorial design, because even for moderately large values of k, the total number of observations n would have to be large to get integer np_i's. For linear models, the accepted practice is to use regular fractions due to the many desirable properties like minimum aberration and optimality. We will show that in our setup the regular fractions are often not optimal. We start by identifying situations when they are optimal.

We use 2^3 designs for illustration. The model matrix for 2^3 main-effects model consists of the first four columns of X given in (2.1) and w_j represents the information in the jth experimental condition, the jth row of X. Suppose the maximum number of experimental conditions is fixed at a number less than 8, and the problem is to identify the experimental conditions and corresponding p_i's that optimize the objective function. Half fractions use 4 experimental conditions (hence the design is uniform). The half fractions defined by rows $\{1, 4, 6, 7\}$ and $\{2, 3, 5, 8\}$ are regular fractions, given by the defining relations $1 = ABC$ and $-1 = ABC$ respectively. If all regression coefficients except the intercept are zeros, then the regular fractions are D-optimal, since all the w_i's are equal. The following theorem identifies the necessary and sufficient conditions for regular fractions to be D-optimal in terms of w_i's.

Theorem 4. *For the 2^3 main-effects model, suppose $\beta_1 = 0$. The regular fractions $\{1, 4, 6, 7\}$ and $\{2, 3, 5, 8\}$ are D-optimal within the class of half-fractions if and only if*

$$4 \ \min\{w_1, w_2, w_3, w_4\} \geq \max\{w_1, w_2, w_3, w_4\}.$$

If $\beta_1 = \beta_2 = 0$, the two regular half-fractions $\{1, 4, 6, 7\}$ and $\{2, 3, 5, 8\}$ are D-optimal half-fractions if and only if $4 \min\{w_1, w_2\} \geq \max\{w_1, w_2\}$.

Example 2. Under the logit link, consider the 2^3 main-effects model with $\beta_1 = \beta_2 = 0$, implying $w_1 = w_3 = w_5 = w_7$ and $w_2 = w_4 = w_6 = w_8$. The regular half-fractions $\{1, 4, 6, 7\}$ and $\{2, 3, 5, 8\}$ have the same $|X'WX|$ but not the same $X'WX$. They are D-optimal half-fractions if and only if one of the following holds:

$$\text{(i) } |\beta_3| \leq \log 2, \tag{4.1}$$

$$\text{(ii) } |\beta_3| > \log 2 \quad \text{and} \quad |\beta_0| \leq \log\left(\frac{2e^{|\beta_3|} - 1}{e^{|\beta_3|} - 2}\right).$$

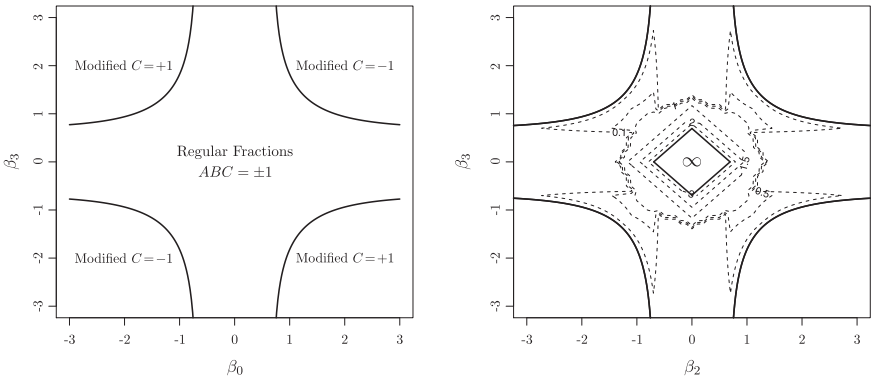

Figure 2. Partitioning of the parameter space.

When the regular half-fractions are not optimal, it follows from Lemma 1 that the goal is to find $\{i_1, i_2, i_3, i_4\}$ that maximizes $|X[i_1, i_2, i_3, i_4]|^2 w_{i_1} w_{i_2} w_{i_3} w_{i_4}$. In this case there are only two distinct w_i's. If $\beta_0\beta_3 > 0$, w_i's corresponding to $\{2, 4, 6, 8\}$ are larger than others, so the fraction given by $C = -1$ will maximize $w_{i_1} w_{i_2} w_{i_3} w_{i_4}$. But this leads to a singular model matrix. It is not surprising that the D-optimal half-fractions are "close" to the design $\{2, 4, 6, 8\}$, and are in fact given by the 16 designs consisting of three elements from $\{2, 4, 6, 8\}$ and one from $\{1, 3, 5, 7\}$. We call these *modified* $C = -1$ fractions. All 16 designs lead to the same $|X'WX|$, $w_1 w_2^3/4$. For $\beta_0\beta_3 < 0$, D-optimal half-fractions are similarly obtained from the fraction $C = +1$.

Figure 2 partitions the parameter space for the 2^3 main-effects logit model. The left panel corresponds to the case (a) $\beta_1 = \beta_2 = 0$. Here the parameters in the middle region would make the regular fractions D-optimal, whereas the top-right and bottom-left regions correspond to the case $\beta_0\beta_3 > 0$. Similarly the other two regions correspond to the case $\beta_0\beta_3 < 0$ so that modified $C = -1$ is optimal. The right panel of Figure 2 is for the case (b) $\beta_1 = 0$ and shows the contour plots for the largest $|\beta_0|$'s that would make the regular fractions D-optimal. (For details, see the Supplementary Materials of this paper.) Along with Figure 2, conditions (4.1) and (S.1) in the Supplementary Materials indicate that if β_1,β_2, and β_3 are small then regular fractions are preferred (see also Table 3). However, when at least one $|\beta_i|$ is large, the regular fractions may not be optimal.

In general, when all the β_i's are nonzero, the regular fractions given by the rows $\{1, 4, 6, 7\}$ or $\{2, 3, 5, 8\}$ are not necessarily the optimal half-fractions. To explore this, we simulated the regression coefficients β_0, β_1, β_2, β_3 independently from different distributions and calculated the corresponding $\mathbf{w}$'s under logit, probit, and complementary log-log links 10,000 times each. For each $\mathbf{w}$, we found the best (according to D-criterion) design supported on 4 distinct rows of

the model matrix. By Lemma 1, any such design has to be uniform. Table 3 gives the percentages of times each of those designs turn out to be the optimal ones for the logit model (the results are somewhat similar for the other links). This shows that the regular fractions are optimal when the β_i's are close to zero. In Table 3, we only report the non-regular fractions which turn out to be D-optimal more than 15% of the times. For the 2^4 case, similarly, when the β_i's are nonzeros, the performance of the regular fractions given by $1 = \pm ABCD$ are not very efficient in general.

We have done a simulation study to determine the efficiency of fractions, especially the regular ones. In order to describe a measure of efficiency, denote the D-criterion value as $\psi(\mathbf{p}, \mathbf{w}) = |X'WX|$ for given $\mathbf{w} = (w_1, \ldots, w_{2^k})'$ and $\mathbf{p} = (p_1, \ldots, p_{2^k})'$. Suppose $\mathbf{p}_w$ is a D-optimal allocation with respect to $\mathbf{w}$. Then the loss of efficiency of $\mathbf{p}$ (with respect to a D-optimal allocation $\mathbf{p}_w$) given $\mathbf{w}$ can be defined as

$$R(\mathbf{p}, \mathbf{w}) = 1 - \left(\frac{\psi(\mathbf{p}, \mathbf{w})}{\psi(\mathbf{p}_w, \mathbf{w})} \right)^{1/(d+1)}. \tag{4.2}$$

In Table 3, we provide within parentheses (the first number) the percentages of times that the regular fractions are at least 70% efficient compared to the best half-fractions (it would correspond to the case where 42% more runs are needed due to a poor choice of design). The second number within the parentheses is the median efficiency. It is clear that when the regular fractions are not D-optimal, they are usually not highly efficient either.

Remark 4. For each of the five situations described in Table 3, we also calculated the corresponding EW D-optimal half-fractions. For all five cases including the highly asymmetric fifth scenario, the regular fractions were EW D-optimal half-fractions.

Remark 5. In Table 2 (and later Table 3 and Table 6) we have used distributions for $\boldsymbol{\beta}$ in two ways. For locally D-optimal designs these distributions were used to simulate the assumed values in order to study the properties of the designs, especially robustness. For EW D-optimal designs these distributions were used as priors.

Remark 6. The priors for $\boldsymbol{\beta}$ should be chosen carefully in applications. A uniform prior on $\beta_i \sim [-a, a]$ can be used when the experimenter does not know much about the corresponding factor. The prior $\beta_i \sim [0, b]$ can be used when the experimenter knows the direction of the corresponding factor effect. In our odor study example, factor A (algae) has two levels: raffinated or solvent extracted algae (-1) and catfish pond algae ($+1$). The scientists initially assessed that

Table 3. Distribution of D-optimal half-fractions under 2^3 main-effects model.

Rows		Percentages				
	$\beta_0 \sim$		$U(-10,10)$			$N(0,5)$
Simulation	$\beta_1 \sim$	$U(-.3,.3)$	$U(-3,3)$	$U(-3,0)$	$U(0,1)$	$N(1,1)$
Setup	$\beta_2 \sim$	$U(-.3,.3)$	$U(0,3)$	$U(0,3)$	$U(0,3)$	$N(2,1)$
	$\beta_3 \sim$	$U(-.3,.3)$	$U(1,5)$	$U(-2,2)$	$U(0,5)$	$N(3,1)$
1467		47.89	0.07	0.86	0.95	0.04
		(100,99.9)	(1.6,15.0)	(8.7,29.2)	(8.8,25.9)	(1.7,18.7)
2358		42.02	0.04	0.68	1.04	0.08
		(100,99.9)	(1.6,15.2)	(8.9,29.1)	(8.7,25.9)	(1.8,18.6)
1235			16.78		35.62	21.50
1347				19.98		
1567			17.45	19.21		
2348			17.54	19.11		
2568				20.01		
4678			16.12		35.41	21.65

raffinated algae has residual lipid which should prevent absorber to interact with volatiles, causing odor to release. Hence it is expected that β_i for this factor should be nonnegative. In this case, one might take the prior on $[0, b]$. For factor B (Scavenger), it is not known before conducting the experiment whether Activated Carbon (-1) is better or worse than Zeolite $(+1)$. Here a symmetric prior on $[-a, a]$ would be more appropriate.

Remark 7. Consider the problem of obtaining the locally D-optimal fractional factorial designs when the number of experimental settings (m, say) is fixed. If the total number of factors under consideration is not too large, one can always calculate the D-efficiencies of all fractions and choose the best one. However, this is a computationally expensive strategy for large k's so we need an alternative. One such strategy would be to choose the m largest w_i's and the corresponding rows, since those w_i represent the information at the corresponding design points. Another one would be to use our algorithms discussed in Section 3.3 to find an optimal allocation for the full factorial designs first, then to choose the m largest p_i's and scale them appropriately. One has to be careful, however, in order to avoid designs which would not allow the estimation of the model parameters. In this case, the exchange algorithm described in Section 3.3.2 can be used to choose the fraction with given m experimental units. Our simulations (not presented here) show that both of these methods perform satisfactorily with the second method giving designs that are generally more than 95% efficient for four factors with the main-effects model. This method is used for computations in the next section.

5. Robustness

5.1. Most robust minimally supported designs

Minimally supported designs have been studied extensively. For continuous or quantitative factors, these designs can be D-optimal for many linear and non-linear models. In our setup of qualitative factors, these designs are attractive since they use the minimal number, $d+1$, of experimental conditions. In many applications, fewer experimental conditions are desirable. In this section, we examine the robustness of minimally supported designs. Our next result gives necessary and sufficient conditions for a fraction to be a D-optimal minimally supported design; it is an immediate consequence of Lemma 1.

Theorem 5. *Let* $\mathbf{I} = \{i_1, \ldots, i_{d+1}\} \subset \{1, \ldots, 2^k\}$ *be an index set. A design* $\mathbf{p}_I = (p_1, \ldots, p_{2^k})'$ *satisfying* $p_i = 0, \forall i \notin I$ *is D-optimal among minimally supported designs if and only if*

$$p_{i_1} = \cdots = p_{i_{d+1}} = \frac{1}{d+1} \text{ and } \mathbf{I} \text{ maximizes } |X[i_1, \ldots, i_{d+1}]|^2 w_{i_1} \cdots w_{i_{d+1}} .$$

For investigating the robustness of a design, we define the maximum loss of efficiency of a given design $\mathbf{p}$ with respect to a specified region $\mathcal{W}$ of $\mathbf{w}$ by

$$R_{\max}(\mathbf{p}) = \max_{\mathbf{w} \in \mathcal{W}} R(\mathbf{p}, \mathbf{w}). \tag{5.1}$$

It can be shown that the region $\mathcal{W}$ takes the form of $[a,\ b]^{2^k}$ for 2^k main-effects model if the range of each of the regression coefficients is an interval symmetric about 0. For example, for a 2^4 main-effects model, if all the regression coefficients range between $[-3, 3]$, then $\mathcal{W} = [3.06 \times 10^{-7},\ 0.25]^{16}$ for logit link, and $[8.33 \times 10^{-49}, 0.637]^{16}$ for probit link. This is the rationale for the choice of the range of w_i's in Theorem 6 below. A design that minimizes the maximum loss of efficiency is called *most robust*. This criterion is also known as "maximin efficiency" in the literature (see, for example, Dette (1997)). For unbounded β_i's with a prior distribution, one can use a quantile instead of the maximum loss to measure robustness.

Theorem 6. *Suppose* $k \geq 3$, $d(d+1) \leq 2^{k+1} - 4$, *and that* $w_i \in [a,\ b]$, $i = 1, \ldots, 2^k$, $0 < a < b$. *Let* $\mathbf{I} = \{i_1, \ldots, i_{d+1}\}$ *be an index set that maximizes* $|X[i_1, i_2, \ldots, i_{d+1}]|^2$. *Then the design* $\mathbf{p}_I = (p_1, \ldots, p_{2^k})'$ *satisfying* $p_{i_1} = \cdots = p_{i_{d+1}} = 1/(d+1)$ *is a most robust minimally supported design with maximum loss* $1 - a/b$ *in efficiency compared to other minimally supported designs.*

Based on Theorem 6, the maximum loss of efficiency depends on the range of w_i's. The result is meaningful only if the interval $[a, b]$ is bounded away from

0. Figure 3 provides some idea about the possible bounds of w_i's for commonly used link functions. For example, for 2^3 designs with the main-effects model, if $0.105 \le w_i \le 0.25$ under the logit link (see Remark 4.1.1 of Yang, Mandal, and Majumdar (2012)), then the maximum loss of efficiency of the regular half-fractional design satisfying $p_1 = p_4 = p_6 = p_7 = 1/4$ is $1 - 0.105/0.25 = 58\%$. The more certain we are about the range of w_i's, the more useful the result is.

For $k = 2$, all 4 minimally supported designs perform equally well (or equally badly). So they are all most robust under our definition. For main-effects models, the condition $d(d+1) \le 2^{k+1} - 4$ in Theorem 6 is guaranteed whenever $k \ge 3$. A most robust minimally supported design can be obtained by searching for an index set $\{i_1, \ldots, i_{d+1}\}$ that maximizes $|X[i_1, i_2, \ldots, i_{d+1}]|^2$. Such an index set is usually not unique. Based on Lemma S1.4, if the index set $\{i_1, \ldots, i_{d+1}\}$ maximizes $|X[i_1, \ldots, i_{d+1}]|^2$, then there always exists another index set $\{i'_1, \ldots, i'_{d+1}\}$ such that $|X[i_1, \ldots, i_{d+1}]|^2 = |X[i'_1, \ldots, i'_{d+1}]|^2$. A most robust minimally supported design may involve a set of experimental conditions $\{i_1, \ldots, i_{d+1}\}$ which does not maximize $|X[i_1, \ldots, i_{d+1}]|^2$. For example, consider a 2^{3-1} design with main-effects model. Suppose $w_i \in [a, b]$, $i = 1, \ldots, 8$. If $4a > b$, then the most robust minimally supported designs are the 2^{3-1} regular fractions. Otherwise, if $4a \le b$, any uniform design restricted to $\{i_1, i_2, i_3, i_4\}$ satisfying $|X[i_1, i_2, i_3, i_4]| \ne 0$ is a most robust minimally supported design.

5.2. Robustness of uniform designs

Yang, Mandal, and Majumdar (2012) showed that for a 2^2 main-effects model, the uniform design is the most robust design in terms of maximum loss of efficiency. In this section, we use simulation studies to examine the robustness of uniform designs and EW D-optimal designs for higher order cases.

For illustration, we use a 2^4 main-effects model. We simulated $\beta_0, \ldots, \beta_4$ from different distributions 1,000 times each and calculated the corresponding $\mathbf{w}$'s, denoted by vectors $\mathbf{w}_1$, ..., $\mathbf{w}_{1000}$. For each $\mathbf{w}_s$, we use the algorithm described in Section 5.1 to obtain a D-optimal allocation $\mathbf{p}_s$. For any allocation $\mathbf{p}$, let $R_{100\alpha}(\mathbf{p})$ be the αth quantile of the set of loss of efficiencies $\{R(\mathbf{p}, \mathbf{w}_s),\ s = 1, \ldots, 1,000\}$. Thus $R_{100}(\mathbf{p}) = R_{\max}(\mathbf{p})$, defined in (5.1) with $\mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_{1000}\}$. The quantities $R_{99}(\mathbf{p})$ and $R_{95}(\mathbf{p})$ are more reliable in measuring the robustness of $\mathbf{p}$.

Table 4 compares $R_{100\alpha}$ for the uniform design $\mathbf{p}_u = (1/16, \ldots, 1/16)'$ with the minimum of $R_{100\alpha}(\mathbf{p}_s)$ for the optimal allocations $\mathbf{p}_s$, $s = 1, \ldots, 1,000$, as well as the $R_{100\alpha}$ of the EW design $\mathbf{p}_e$. In this table, if the values of column (I) are smaller than those of column (II), then we can conclude that the uniform design is better than all the D-optimal designs in terms of the quantiles of loss of efficiency. This happens in many situations. Table 4 provides strong evidence

Table 4. Loss of efficiency of 2^4 uniform design.

	Percentages											
	$\beta_0 \sim U(-3,3)$			$U(-1,1)$			$U(-3,0)$			$N(0,5)$		
	$\beta_1 \sim U(-1,1)$			$U(0,1)$			$U(1,3)$			$N(0,1)$		
Simulation	$\beta_2 \sim U(-1,1)$			$U(0,1)$			$U(1,3)$			$N(2,1)$		
Setup	$\beta_3 \sim U(-1,1)$			$U(0,1)$			$U(-3,-1)$			$N(-.5,2)$		
	$\beta_4 \sim U(-1,1)$			$U(0,1)$			$U(-3,-1)$			$N(-.5,2)$		
Quantiles	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)
R_{99}	0.348	0.353	0.348	0.146	0.111	0.112	0.503	0.273	0.299	0.650	0.864	0.726
R_{95}	0.299	0.304	0.299	0.128	0.094	0.093	0.495	0.251	0.256	0.617	0.788	0.670
R_{90}	0.271	0.274	0.271	0.117	0.084	0.085	0.488	0.239	0.233	0.589	0.739	0.629

Note: (I) $= R_{100\alpha}(\mathbf{p}_u)$, (II) $= \min_{1\le s\le 1,000} R_{100\alpha}(\mathbf{p}_s)$, (III) $= R_{100\alpha}(\mathbf{p}_e)$.
$\mathbf{p}_u$ is the uniform design, $\mathbf{p}_s$ is the locally D-optimal design and $\mathbf{p}_e$ is the EW D-optimal design.

that the uniform design $\mathbf{p}_u$ is one of the most robust ones if the β_i's are expected to come from an interval that is symmetric around zero. This is consistent with the conclusion of Cox (1988).

However, there are situations where the uniform design does not perform well, as illustrated by the two middle blocks of Table 4. If the signs of the regression coefficients are known, it is advisable not to use the uniform design. For many practical applications, the experimenter will have some idea of the direction of effects of factors, which in statistical terms determines the signs of the regression coefficients. For these situations, it turns out that the performance of the EW D-optimal designs is comparable to that of the most robust designs, even when the uniform design does not perform well (see columns (III) in Table 4, where $\mathbf{p}_e$ is the EW design). Hence we recommend the use of EW D-optimal designs when the experimenter has some idea about the signs of β_i's. Uniform designs are recommended in the absence of prior knowledge of the sign of the regression parameters.

Consider the uniform designs restricted to regular fractions. Again we use 2^4 main-effects model as illustration and consider the uniform designs restricted to the regular half-fractions identified by $1 = \pm ABCD$. We performed simulations as above and our conclusions are similar: uniform designs on regular fractions are among the most robust ones if the signs of the regression parameters are unknown but they may not perform well if the signs of β_i's are known.

6. Examples

Example 3. We revisit the odor examples discussed in the introduction. The 2^{4-1}_{IV} design given by $D = -ABC$ was used with 5 replications per experimental setup. For factor C, the polypropylene used in this experiment is in tiny crystal

Table 5. Optimal design for the Odor Study.

A	B	C	D	E(w_i)	$\mathbf{n}_{odor}$	$\mathbf{n}_{EW}$	$\mathbf{n}_{EW1/2}$
+1	+1	+1	+1	0.050			
+1	+1	+1	−1	0.105	5	3	7
+1	+1	−1	+1	0.105	5	4	3
+1	+1	−1	−1	0.105		3	
+1	−1	+1	+1	0.050	5		
+1	−1	+1	−1	0.105		4	
+1	−1	−1	+1	0.105		3	4
+1	−1	−1	−1	0.105	5	3	6
−1	+1	+1	+1	0.105	5	4	
−1	+1	+1	−1	0.105		3	3
−1	+1	−1	+1	0.105		2	7
−1	+1	−1	−1	0.050	5	1	
−1	−1	+1	+1	0.105		3	6
−1	−1	+1	−1	0.105	5	3	4
−1	−1	−1	+1	0.105	5	4	
−1	−1	−1	−1	0.050			

form as opposed to fine powder which leads the scientist to speculate that β_3 should be positive. Moreover one expects that the presence of compatabilizers should reduce the odor and hence β_4 is expected to be positive. Initial results from the experiment indicate that the number of successes is increasing in the level of A (from -1 to $+1$). We examine the efficiency of the design used in this experiment in view of these facts and consider an EW D-optimal design with the ranges $(-3, 3)$ for β_0, β_2 and (0,3) for $\beta_1, \beta_3, \beta_4$. These priors are reasonably uninformative except for the directions of effects of the factors (signs of the parameters). Furthermore, if the design points are not restricted to the original half-fraction, the best EW D-optimal design with 40 experimental units, given by $\mathbf{n}_{EW}$, is supported on 13 points.

In order to compare the performance of the three designs given in Table 5, we drew 1,000 random samples of the β_i's from the setup discussed above and for each of them calculated the locally D-optimal design with 40 runs. Then we calculated the loss of efficiencies of the EW D-optimal design ($\mathbf{n}_{EW}$) and EW D-optimal half-fraction ($\mathbf{n}_{EW\frac{1}{2}}$) as well as that of the original design used ($\mathbf{n}_{odor}$), with respect to the locally D-optimal design. The mean, standard deviation, and some quantiles of the loss of efficiencies are given in Table 6. These numbers indicate that the EW D-optimal design is around 20% more efficient than the original one, while the EW half-fraction design is about 10% more efficient than the original one.

Example 4. Hamada and Nelder (1997) discussed a 2^{4-1} fractional factorial experiment performed at IIT Thompson laboratory that was originally reported

Table 6. Odor Study: Loss of efficiencies of different designs.

Design	R_{99}	R_{95}	R_{90}	Mean	SD
EW design ($\mathbf{n}_{EW}$)	51.4	46.6	44.7	33.0	9.5
EW half-fraction ($\mathbf{n}_{EW\frac{1}{2}}$)	77.2	69.5	63.2	41.9	15.7
Original design ($\mathbf{n}_{odor}$)	84.8	76.8	70.1	51.8	15.1

by Martin, Parker, and Zenick (1987). This was a windshield molding slugging experiment where the outcome was whether the molding was good or not. There were four factors each at two levels: (A) poly-film thickness (0.0025, 0.00175), (B) oil mixture ratio (1:20, 1:10), (C) material of gloves (cotton, nylon), and (D) the condition of metal blanks (dry underside, oily underside). By analyzing the data presented in Hamada and Nelder (1997), we get an estimate of the unknown parameter as $\hat{\boldsymbol{\beta}} = (1.77, -1.57, 0.13, -0.80, -0.14)'$ under logit link. If one wants to conduct a follow-up experiment on half-fractions, then it is sensible to use the knowledge obtained by analyzing the data. With the knowledge of $\hat{\boldsymbol{\beta}}$, we take the assumed value of $\boldsymbol{\beta}$ as $(2, -1.5, 0.1, -1, -0.1)'$. The locally D-optimal design $\mathbf{p}_a$ is given in Table 7. Another option is to consider a range for the possible values of the regression parameters, namely, $(1, 3)$ for β_0, $(-3, -1)$ for β_1, $(-0.5, 0.5)$ for β_2, β_4, and $(-1, 0)$ for β_3. For this choice of range for the parameter values with independence and uniform distributions, the EW D-optimal half-fractional design $\mathbf{p}_e$ is also given in Table 7. We have calculated the linear predictor η and success probability π for all possible experimental settings. It seems that a good fraction would not favor high success probabilities very much. This is one of the main differences between the design reported by Hamada and Nelder (denoted by $\mathbf{p}_{HN}$) and our designs (denoted by $\mathbf{p}_a$ and $\mathbf{p}_e$). These designs have six rows in common. The last two columns of Table 7 give the Baysian D-optimal and EW D-optimal designs, respectively. It can be seen that the optimal allocations for these two designs are quite similar, and both of them are supported on the same rows.

7. Discussion and Future Research

For binary response, the logit link is the most commonly used link in practice. The situation under this link function is close to that in the linear model case because, typically, the w_i's are not too close to 0 and do not vary much. Similar to the cases of linear models, uniform designs perform well under logit link, more than other popular link functions. In general, the performance of the logit and probit links are similar, while that of the complementary log-log link is somewhat different. For example, if we consider a 2^2 experiment with a main-effects model, the efficiency of the uniform design with respect to the Bayes D-optimal design is 99.99% under the logit link, but is only 89.6% under the

Table 7. Optimal half-fraction design for Windshield Molding Experiment.

Row	A	B	C	D	η	π	$\mathbf{p}_{HN}$	$\mathbf{p}_a$	$\mathbf{p}_e$	$\mathbf{p}_B$	$\mathbf{p}_{e_f}$
5	+1	−1	+1	+1	-0.87	0.295		0.044	0.184	0.073	0.092
1	+1	+1	+1	+1	-0.61	0.352	0.125	0.178	0.011	0.117	0.103
6	+1	−1	+1	−1	-0.59	0.357	0.125	0.178	0.011	0.118	0.103
2	+1	+1	+1	−1	-0.33	0.418		0.059	0.184	0.078	0.092
7	+1	−1	−1	+1	0.73	0.675	0.125	0.163		0.125	0.103
3	+1	+1	−1	+1	0.99	0.729			0.195	0.079	0.091
8	+1	−1	−1	−1	1.01	0.733			0.195	0.078	0.091
4	+1	+1	−1	−1	1.27	0.781	0.125	0.147		0.115	0.103
13	−1	−1	+1	+1	2.27	0.906	0.125	0.158	0.111	0.061	0.054
9	−1	+1	+1	+1	2.53	0.926				0.053	0.057
14	−1	−1	+1	−1	2.55	0.928				0.043	0.057
10	−1	+1	+1	−1	2.81	0.943	0.125	0.074	0.110	0.061	0.053
15	−1	−1	−1	+1	3.87	0.980					
11	−1	+1	−1	+1	4.13	0.984	0.125				
16	−1	−1	−1	−1	4.15	0.984	0.125				
12	−1	+1	−1	−1	4.41	0.988					

Notation: $\mathbf{p}_{HN}$: Design reported by Hamada and Nelder, $\mathbf{p}_a$: Locally D-optimal design, $\mathbf{p}_e$: EW D-optimal half-fraction, $\mathbf{p}_B$: Bayesian D-optimal design, $\mathbf{p}_{e_f}$: EW D-optimal design.

complementary log-log link. Figure 3 provides a graphical display of the weight function (w) for commonly used link functions. As seen there, the complementary log-log link function is not symmetric about 0. This partly explains the poor performance of the uniform design under this link. Nevertheless, the EW D-optimal designs are still highly efficient across different link functions. For the same setup, the efficiencies of EW designs with respect to the corresponding Bayesian D-optimal designs are 99.99% (logit link), 99.94% (probit link), 99.77% (log-log link), and 100.00% (complementary log-log link), respectively. It appears that EW D-optimal designs are excellent surrogates of Bayes D-optimal designs. A more extensive investigation is planned for the future.

Efficiencies depend on the priors used for the parameters, and hence the prior on the $\boldsymbol{\beta}$s should be different for different link functions in order to maintain roughly consistent prior beliefs about the success probabilities under different experimental setups.

Our recommendation is to use EW D-optimal designs unless the experimenter has absolutely no prior knowledge of the parameters, in which case it is recommended to use the uniform design. In EW optimality, we replace the w_i's by their expectations. However, taking the average of the w_i's is not same as taking the average of the β_i's. Consider a 2^4 design with main-effects model. Table 8 uses the notations from Table 4. Suppose $\beta_0 \sim U(-3, 0)$, $\beta_1, \beta_2 \sim U(1, 3)$, $\beta_3, \beta_4 \sim U(-3, -1)$, and the β_i's are independent. It is clear that the uniform

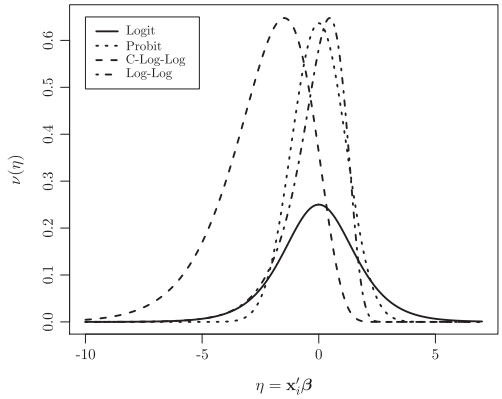

Figure 3. $w_i = \nu(\eta_i) = \nu(\mathbf{x}_i'\boldsymbol{\beta})$ for commonly used link functions.

Table 8. Loss of efficiencies of different designs for 2^4 main-effects model.

	Uniform	Most robust	EW D-opt	$E(\beta)$ D-opt
R_{99}	0.503	0.273	0.299	0.331
R_{95}	0.495	0.251	0.256	0.284
R_{90}	0.488	0.239	0.233	0.251

design performs much worse compared to the most robust design, while the performance of the EW D-optimal design is comparable with the best design. The last column corresponds to the locally D-optimal design where the assumed values of the parameter are taken to be the midpoints of the ranges of β_i's mentioned above. Clearly this is worse than the EW D-optimal design.

In the linear model setup, as the potential columns in the model matrix are orthogonal, analysis of experimental data based on regular fractions is not unduly biased by the omission of non-negligible model terms. Under a GLM setup, the regular fractions may give larger than necessary variance for some models. We did not consider the performance of different designs under model robustness. Moreover, because of the bias-variance trade-off, regular fractions (or other designs) may not be model-robust. Extending optimal designs based on GLMs to topics such as confounding, aberration, and trade-off between variance and bias represents an important topic for future research.

Supplementary Materials

The proofs of the Theorems 2, 3, 4, 6 and some associated lemmas in this paper are given in the online supplementary material available at `http://www3.stat.sinica.edu.tw/statistica/`. There is also a discussion of the connection between general equivalence theorem and Theorem 1 and some additional results for Example 4.1, as well as a discussion on the exchange algorithm for real-valued allocations.

Acknowledgement

We thank Dr. Suraj Sharma for providing the details of the Odor Study, and Dr. Yaming Yu for sharing codes for the Multiplicative and Cocktail algorithms. We also thank the reviewers for comments and suggestions that substantially improved the quality of the manuscript. This research is in part supported by NSF Grant DMS-09-05731 and NSA Grant H98230-13-1-0251.

References

Agresti, A. (2013). *Categorical Data Analysis*. 3rd edition. Wiley, New York.

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

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: a review. *Statist. Sci.* **10**, 273-304.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.

Cox, D. R. (1988). A note on design when response has an exponential family distribution. *Biometrika* **75**, 161-164.

Dette, H. (1997). Designing experiments with respect to 'standardized' optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59**, 97-110.

Dobson, A. J. and Barnett, A. (2008). *An Introduction to Generalized Linear Models*. 3rd edition. Chapman and Hall/CRC, London.

Dorta-Guerra, R., González-Dávila, E. and Ginebra, J. (2008). Two-level experiments for binary response data. *Comput. Statist. Data Anal.* **53**, 196-208.

Dror, H. A. and Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520-529.

Dror, H. A. and Steinberg, D. M. (2008). Sequential experimental designs for generalized linear models. *J. Amer. Statist. Assoc.* **103**, 288-298.

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

Fedorov, V. V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*. Springer, New York.

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

Ford, I., Titterington, D. M. and Kitsos, C. P. (1989). Recent advances in nonlinear experimental design. *Technometrics* **31**, 49-60.

González-Dávila, E., Dorta-Guerra, R. and Ginebra, J. (2007). On the information in two-level experiments. *Model Assisted Statist. and Appl.* **2**, 173-187.

Gotwalt, C. M., Jones, B. A. and Steinberg, D. M. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics* **51**, 88-95.

Graßhoff, U. and Schwabe, R. (2008). Optimal design for the Bradley-Terry paired comparison model. *Statist. Methods and Appl.* **17**, 275-289.

Hamada, M. and Nelder, J. A. (1997). Generalized linear models for quality-improvement experiments. *J. Quality Tech.* **29**, 292-304.

Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Ann. Statist.* **29**, 561-576.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849-879.

Khuri, A. I., Mukherjee, B., Sinha, B. K. and Ghosh, M. (2006). Design issues for generalized linear models: A Review. *Statist. Sci.* **21**, 376-399.

Li, G. and Majumdar, D. (2008). D-optimal designs for logistic models with three and four parameters. *J. Statist. Plann. Inference* **138**, 1950-1959.

Li, G. and Majumdar, D. (2009). Some results on D-optimal designs for nonlinear models with applications. *Biometrika* **96**, 487-493.

Lindsey, J. (1997). *Applying Generalized Linear Models*. Springer, New York.

Mandal, A., Wong, W. K. and Yu, Y. (2015). Algorithmic searches for optimal designs. In *Handbook of Design and Analysis of Experiments*, (Edited by A. Dean, M. Morris, J. Stufken, and D. Bingham), Chapman and Hall/CRC.

Martin, B., Parker, D. and Zenick, L. (1987). Minimize slugging by optimizing controllable factors on topaz windshield molding. In *Fifth Symposium on Taguchi Methods*, 519-526. American Supplier Institute, Inc., Dearborn, MI.

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

McCulloch, C. and Searle, S. (2001). *Generalized, Linear and Mixed Models*. Wiley, New York.

Müller, W. G. (2007). *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*, 3rd Edition, Springer.

Myers, R. M., Montgomery, D. C. and Vining, G. G. (2002). *Generalized Linear Models with Applications in Engineering and Statistics*. Wiley, New York.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer, New York.

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

Pronzato, L. and Walter, E. (1988). Robust experiment design via maximin optimization. *Math. Biosci.* **89**, 161-176.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.

Russell, K. G., Woods, D. C., Lewis, S. M. and Eccleston, J. A. (2009). D-optimal designs for Poisson regression models. *Statist. Sinica* **19**, 721-730.

Silvey, S. D., Titterington, D. M. and Torsney, B. (1978). An algorithm for optimal designs on a finite design space. *Comm. Statist. Theory Methods* **14**, 1379-1389.

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). Wiley, New York.

Titterington, D. M. (1976). Algorithms for computing D-optimal design on finite design spaces. In *Proc. of the 1976 Conf. on Information Science and Systems*, **3**, 213-216. John Hopkins University.

Titterington, D. M. (1978). Estimation of correlation coefficients by ellipsoidal trimming. *Appl. Stat.* **27**, 227-234.

Waterhouse, T. H., Woods, D. C., Eccleston, J. A. and Lewis, S. M. (2008). Design selection criteria for discrimination/estimation for nested models and a binomial response. *J. Statist. Plann. Inference* **138**, 132-144.

Woods, D. C., Lewis, S. M., Eccleston, J. A. and Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48**, 284-292.

Woods, D. C. and van de Ven, P. (2011). Blocked designs for experiments with non-normal response. *Technometrics* **53**, 173-182.

Wynn, H. P. (1970). The sequential generation of D-optimum experimental designs. *Ann. Math. Statist.* **41**, 1655-1664.

Xu, H., Phoa, F. K. H. and Wong, W. K. (2009). Recent developments in nonregular fractional factorial designs. *Statist. Surveys* **3**, 18-46.

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

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* **37**, 518-541.

Yang, M., Zhang, B. and Huang, S. (2011). Optimal designs for generalized linear models with multiple design variables. *Statist. Sinica* **21**, 1415-1430.

Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *Ann. Statist.* **38**, 1593-1606.

Zangwill, W. (1969). *Nonlinear Programming: A Unified Approach*. Prentice-Hall, New Jersey.

Zayats, N. and Steinberg, D. M. (2010). Optimal design of experiments when factors affect detection capability. *Pakistan J. Statist.* **26**, 15-37.

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7045, USA.

E-mail: jyang06@math.uic.edu

Department of Statistics, University of Georgia, Athens, GA 30602, USA.

E-mail: amandal@stat.uga.edu

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7045, USA.

E-mail: dibyen@uic.edu

(Received September 2013; accepted February 2015)