Statistica Sinica **27** (2017), 1879-1902
doi:https://doi.org/10.5705/ss.202016.0210

# D-OPTIMAL DESIGNS WITH ORDERED CATEGORICAL DATA

Jie Yang, Liping Tong and Abhyuday Mandal

*University of Illinois at Chicago, Advocate Health Care*
*and University of Georgia*

*Abstract:* Cumulative link models have been widely used for ordered categorical responses. Uniform allocation of experimental units is commonly used in practice, but often suffers from a lack of efficiency. We consider D-optimal designs with ordered categorical responses and cumulative link models. For a predetermined set of design points, we derive the necessary and sufficient conditions for an allocation to be locally D-optimal and develop efficient algorithms for obtaining approximate and exact designs. We prove that the number of support points in a minimally supported design only depends on the number of predictors, which can be much less than the number of parameters in the model. We show that a D-optimal minimally supported allocation in this case is usually not uniform on its support points. In addition, we provide EW D-optimal designs as a highly efficient surrogate to Bayesian D-optimal designs. Both of them can be much more robust than uniform designs.

*Key words and phrases:* Approximate design, cumulative link model, exact design, minimally supported design, multinomial response, ordinal data.

## 1. Introduction

In this paper we determine optimal and efficient designs for factorial experiments with qualitative factors and ordered categorical responses, or simply ordinal data. Design of experiment with multinomial response, and ordered categories in particular, is becoming increasingly popular in a rich variety of scientific disciplines, especially when human evaluations are involved (Christensen (2015)). Examples include a wine bitterness study (Randall (1989)), potato pathogen experiments (Omer, Johnson and Rowe (2000)), a radish seedling's damping-off study (Krause, Madden and Hoitink (2001)), a polysilicon deposition study (Wu (2008)), beef cattle research (Osterstock et al. (2010)), and a toxicity study (Agresti (2013)).

This research is motivated by an *odor removal study* conducted by the textile engineers at the University of Georgia. The scientists studied the manufacture

Table 1. Pilot study of odor removal study.

| Experimental setting $i$ | Factor Algae $x_1$ | level Resin $x_2$ | Summarized responses ($Y$, odor) Serious $y_{i1}$ | Medium $y_{i2}$ | No odor $y_{i3}$ |
|---|---|---|---|---|---|
| 1 | $+$ | $+$ | 2 | 6 | 2 |
| 2 | $+$ | $-$ | 7 | 2 | 1 |
| 3 | $-$ | $+$ | 0 | 0 | 10 |
| 4 | $-$ | $-$ | 0 | 2 | 8 |

of bio-plastics containing odorous volatiles, that need to be removed before commercialization. For that purpose, a $2^2$ factorial experiment was conducted using algae and synthetic plastic resin blends. The factors were `types of algae` ($x_1$: raffinated or solvent extracted algae ($-$), catfish pond algae ($+$)) and `synthetic resins` ($x_2$: polyethylene ($-$), polypropylene ($+$)). The response $Y$ had three ordered categories: serious odor ($j = 1$), medium odor ($j = 2$), and almost no odor ($j = 3$). Following traditional factorial design theory, a pilot study with equal numbers (10 in this case) of replicates at each experimental setting was conducted, a *uniform design*. The results are summarized in Table 1, where $y_{ij}$ represents the number of responses falling into the $j$th category under the $i$th experimental setting. As demonstrated later (Section 4), the best design identified by our research could improve the efficiency by 25% with only three experimental settings involved.

For such kind of ordinal response $Y$ with $J$ categories and $d$ predictors $\mathbf{x} = (x_1, \ldots, x_d)^T$, the most popular model in practice was first the *proportional odds model* (also known as *cumulative logit model*, see Liu and Agresti (2005) for a detailed review). McCullagh (1980) extended it to the *cumulative link model* (also known as *ordinal regression model*)

$$g\left(P(Y \leq j \mid \mathbf{x})\right) = \theta_j - \boldsymbol{\beta}^T \mathbf{x}, \ j = 1, \ldots, J-1, \tag{1.1}$$

where $g$ is a general link function, with the proportional odds model as a special case when $g$ is the logit link. Examples include the complementary log-log link for the polysilicon deposition study (see Example 6) and the cauchit link for the toxicity study (see Example 9). We adopt the cumulative link model (1.1).

When there are only two categories ($J = 2$), the cumulative link model (1.1) is essentially a generalized linear model for binary data (McCullagh and Nelder (1989); Dobson and Barnett (2008)). For optimal designs under generalized linear models, there is a growing body of literature (see Khuri et al. (2006), Atkinson, Donev and Tobias (2007), Stufken and Yang (2012), and references therein). In this case, it is known that the minimum number of experimental settings required by a nondegenerate Fisher information matrix is $d+1$, which equals the number of parameters (Fedorov (1972); Yang and Mandal (2015)). A design with the least number of experimental settings, known as a *minimally supported design*, is of practical significance with a specified regression model due to the cost of changing settings. It is also known that the experimental units should be uniformly assigned when a minimally supported design is adopted for binary response, or under a univariate generalized linear model (Yang and Mandal (2015)).

When $J \geq 3$, the cumulative link model is a special case of the multivariate generalized linear model (McCullagh (1980)). The relevant results in the optimal design literature are meagre and restricted to the logit link function (Zocchi and Atkinson (1999); Perevozskaya, Rosenberger and Haines (2003)). Here we obtain theoretical results and efficient algorithms for general link functions and reveal that the optimal designs with $J \geq 3$ are quite different from the cases with $J = 2$. We prove that the minimum number of experimental settings is still $d+1$, but strictly less than the number of parameters $d+J-1$ (Theorems 3 and 4). This counter-intuitive result is due to the multinomial-type responses: from a single experimental setup, the summarized responses have $J-1$ degrees of freedom, requiring fewer distinct experimental settings in a minimally supported design. For the same reason, the allocation of replicates in a minimally supported design is usually not uniform (Section 5), which differs from the traditional factorial design theory.

As with generalized linear models, the information matrix under cumulative link models depends on unknown parameters. Different approaches have been proposed to solve the dependence of optimal designs on unknown parameters, including local optimality (Chernoff (1953)), Bayesian approach (Chaloner and Verdinelli (1995)), a maximin approach (Pronzato and Walter (1988); Imhof (2001)), and a sequential procedure (Ford, Titterington and Kitsos (1989)). As pointed out by Ford, Torsney and Wu (1992), locally optimal designs are not only important when good initial parameters are available from previous experiments, but can also be a benchmark for designs chosen to satisfy experimental constraints. We mainly focus on locally optimal designs. For situations where local values of the parameters are difficult to obtain, but the experimenter has an idea of the range of parameters with or without a prior distribution, we recommend EW optimal designs, where the Fisher information matrix is replaced by its expected values (Atkinson, Donev and Tobias (2007); Yang, Mandal and Majumdar (2016)). We compare Bayesian D-optimal designs (Chaloner and Verdinelli (1995)) with EW D-optimal designs for ordinal data. As a surrogate for Bayesian designs, an EW design is much easier to find and retains high efficiency with respect to Bayesian criterion (Section 6).

Among various optimal design criteria, D-optimality, which maximizes the determinant of Fisher information matrix, is the most frequently used (Zocchi and Atkinson (1999)) and often performs well according to other criteria (Atkinson, Donev and Tobias (2007)). We study D-optimal designs.

In the design literature, one type of experiment deals with quantitative or continuous factors only. Such a design problem includes the identification of a set of design points $\{\mathbf{x}_i\}_{i=1,\ldots,m}$ and the corresponding weights $\{p_i\}_{i=1,\ldots,m}$ (see, for example, Atkinson, Donev and Tobias (2007) and Stufken and Yang (2012)). Numerical algorithms are typically used for cases with two or more factors (see, for example, Woods et al. (2006)). Another type of experiment employs qualitative or discrete factors, where the set of design points $\{\mathbf{x}_i\}_{i=1,\ldots,m}$ is predetermined and only the weights $\{p_i\}_{i=1,\ldots,m}$ are to be optimized (see, for example, Yang and Mandal (2015)). One can pick grid points of continuous factors and turn the first kind of problem into the second. Tong, Volkmer and Yang (2014, Sec. 5) also bridged the gap between the two types of problems in a way that results involving discrete factors can be applied to the cases with continuous factors. We concentrate on the second kind of design problems and assume that $\{\mathbf{x}_i\}_{i=1,\ldots,m}$ are given and fixed.

This paper is organized as follows. In Section 2, we describe the preliminary setup and obtain the Fisher information matrix for the cumulative link model with a general link, generalizing Perevozskaya, Rosenberger and Haines (2003). We also identify a necessary and sufficient condition for the Fisher information matrix to be positive definite, which determines the minimum number of experimental settings required. In Sections 3 and 4, we provide theoretical results and numerical algorithms for searching locally D-optimal approximate or exact designs. In Section 5, we identify analytic D-optimal designs for special cases to illustrate that a D-optimal minimally supported design is usually not uniform on its support points. In Section 6, we illustrate by examples that the EW D-optimal design can be highly efficient with respect to Bayesian D-optimality. We make concluding remarks in Section 7 and relegate additional proofs and results to the supplementary materials.

## 2. Fisher Information Matrix and Its Determinant

Suppose there are $m$ ($m \geq 2$) predetermined experimental settings. For the $i$th experimental setting with corresponding predictors $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T \in \mathbb{R}^d$ ($d \geq 1$), there are $n_i$ experimental units assigned to it. Among the $n_i$ experimental units, the $k$th one generates a response $V_{ik}$ which belongs to one of $J$ ($J \geq 2$) ordered categories. As shown in Example 2, the dimension $d$ of the predictors can be significantly larger than the number of factors considered in the experiment, which allows more flexible models.

### 2.1. General setup

In many applications, $V_{i1}, \ldots, V_{in_i}$ are regarded as i.i.d. discrete random variables. Let $\pi_{ij} = P(V_{ik} = j)$, where $i = 1, \ldots, m$; $j = 1, \ldots, J$; and $k = 1, \ldots, n_i$. Let $Y_{ij} = \#\{k \mid V_{ik} = j\}$ be the number of $V_{ik}$'s falling into the $j$th category. Then $(Y_{i1}, \ldots, Y_{iJ}) \sim \text{Multinomial}(n_i; \pi_{i1}, \ldots, \pi_{iJ})$.

**Assumption 1.** $0 < \pi_{ij} < 1$, $i = 1, \ldots, m$; $j = 1, \ldots, J$.

Let $\gamma_{ij} = P(V_{ik} \leq j) = \pi_{i1} + \cdots + \pi_{ij}$ , $j = 1, \ldots, J$. Based on Assumption 1, $0 < \gamma_{i1} < \gamma_{i2} < \cdots < \gamma_{i,J-1} < \gamma_{iJ} = 1$ for each $i = 1, \ldots, m$. Consider independent multinomial observations $(Y_{i1}, \ldots, Y_{iJ}), i = 1, \ldots, m$ with corresponding predictors $\mathbf{x}_1, \ldots, \mathbf{x}_m$. Under a *cumulative link model* or *ordinal regression model* (McCullagh (1980); Agresti (2013); Christensen (2015)), there exists a link function $g$ and parameters of interest $\theta_1, \ldots, \theta_{J-1}, \boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^T$, such that

$$g(\gamma_{ij}) = \theta_j - \mathbf{x}_i^T \boldsymbol{\beta}, \quad j = 1, \ldots, J-1. \tag{2.1}$$

This leads to $m(J-1)$ equations in $d + J - 1$ parameters $(\beta_1, \ldots, \beta_d, \theta_1, \ldots, \theta_{J-1})$.

**Assumption 2.** *The link $g$ is differentiable and its derivative $g' > 0$.*

Assumption 2 is satisfied for commonly used link functions including `logit` $(\log(\gamma/(1-\gamma)))$, `probit` $(\Phi^{-1}(\gamma))$, `log-log` $(-\log(-\log(\gamma)))$, `complementary log-log` $(\log(-\log(1-\gamma)))$, and `cauchit` $(\tan(\pi(\gamma - 1/2)))$ (McCullagh and Nelder (1989); Christensen (2015)). Some relevant formulas of these link functions are provided in the supplementary materials (Section S.1). According to Assumption 2, $g$ is strictly increasing, and then $\theta_1 < \theta_2 < \cdots < \theta_{J-1}$.

**Example 1.** Consider the logit link $g(\gamma) = \log(\gamma/(1-\gamma))$ with two predictors and three ordered categories. Model (2.1) consists of $2m$ equations $g(\gamma_{ij}) = \theta_j - x_{i1}\beta_1 - x_{i2}\beta_2$, $i = 1, \ldots, m$; $j = 1, 2$ and parameters $(\beta_1, \beta_2, \theta_1, \theta_2)$. Under Assumptions 1 and 2, $\theta_1 < \theta_2$.

**Example 2.** Suppose the model consists of three covariates $x_1, x_2, x_3$ and a few second-order predictors, $g(\gamma_{ij}) = \theta_j - x_{i1}\beta_1 - x_{i2}\beta_2 - x_{i3}\beta_3 - x_{i1}x_{i2}\beta_{12} - x_{i1}^2\beta_{11} - x_{i2}^2\beta_{22}$, where $i = 1, \ldots, m; j = 1, \ldots, J-1$. Then the number of predictors is $d = 6$.

Under the cumulative link model (2.1), the log-likelihood function (up to a constant) is $l(\beta_1, \ldots, \beta_d, \theta_1, \ldots, \theta_{J-1}) = \sum_{i=1}^m \sum_{j=1}^J Y_{ij} \log(\pi_{ij})$, where $\pi_{ij} = \gamma_{ij} - \gamma_{i,j-1}$ with $\gamma_{ij} = g^{-1}(\theta_j - \mathbf{x}_i^T \boldsymbol{\beta})$ for $j = 1, \ldots, J-1$ and $\gamma_{i0} = 0$, $\gamma_{iJ} = 1$.

Perevozskaya, Rosenberger and Haines (2003) obtained a detailed form of the Fisher information matrix for logit link and one predictor. Our result is for general link and $d$ predictors; its proof is relegated to the supplementary materials (Section S.3).

**Theorem 1.** *Under Assumptions* 1 *and* 2*, the Fisher information matrix can be written as*

$$\mathbf{F} = \sum_{i=1}^m n_i \mathbf{A}_i, \tag{2.2}$$

*where* $\mathbf{A}_i$ *is the* $(d+J-1) \times (d+J-1)$ *matrix*

$$\begin{pmatrix} \mathbf{A}_{i1} & \mathbf{A}_{i2} \\ \mathbf{A}_{i2}^T & \mathbf{A}_{i3} \end{pmatrix} = \begin{pmatrix} (e_i x_{is} x_{it})_{s=1,\ldots d; t=1,\ldots,d} & (-x_{is} c_{it})_{s=1,\ldots,d; t=1,\ldots,J-1} \\ (-c_{is} x_{it})_{s=1,\ldots,J-1; t=1,\ldots,d} & \mathbf{A}_{i3} \end{pmatrix}$$

*and* $\mathbf{A}_{i3}$ *is the* $(J-1) \times (J-1)$ *symmetric tri-diagonal matrix with diagonal entries* $u_{i1}, \ldots, u_{i,J-1}$, *and off-diagonal entries* $-b_{i2}, \ldots, -b_{i,J-1}$ *when* $J \geq 3$, *where* $e_i = \sum_{j=1}^J \pi_{ij}^{-1} (g_{ij} - g_{i,j-1})^2 > 0$ *with* $g_{ij} = (g^{-1})'(\theta_j - \mathbf{x}_i^T \boldsymbol{\beta}) > 0$ *for* $j = 1, \ldots, J-1$ *and* $g_{i0} = g_{iJ} = 0$; $c_{it} = g_{it}[\pi_{it}^{-1}(g_{it} - g_{i,t-1}) - \pi_{i,t+1}^{-1}(g_{i,t+1} - g_{it})]$; $u_{it} = g_{it}^2(\pi_{it}^{-1} + \pi_{i,t+1}^{-1}) > 0$; *and* $b_{it} = g_{i,t-1} g_{it} \pi_{it}^{-1} > 0$. $\mathbf{A}_{i3}$ *contains only one entry* $u_{i1}$ *when* $J = 2$.

As the Fisher information matrix, $\mathbf{F}$ is always positive semi-definite, $|\mathbf{F}| \geq 0$ (Fedorov (1972)). As a special case, $\mathbf{A}_i$ is the Fisher information at the experimental setting $\mathbf{x}_i$ (also known as a *design point* or *support point*) and thus is positive semi-definite.

## 2.2. Determinant of Fisher information matrix

Among different criteria for optimal designs, D-criterion looks for the allocation maximizing $|\mathbf{F}|$, the determinant of $\mathbf{F}$. Here, a D-optimal design with $m$ predetermined design points $\mathbf{x}_1, \ldots, \mathbf{x}_m$ could either be an integer-valued allocation $(n_1, n_2, \ldots, n_m)$ maximizing $|\mathbf{F}|$ with fixed $n = \sum_{i=1}^m n_i > 0$, known as an *exact design*; or a real-valued allocation $(p_1, p_2, \ldots, p_m)$ maximizing $|n^{-1}\mathbf{F}|$ with

$p_i = n_i/n \geq 0$ and $\sum_{i=1}^m p_i = 1$, known as an *approximate design*.

**Theorem 2.** *The determinant of the Fisher information matrix,*

$$|\mathbf{F}| = \sum_{\alpha_1+\cdots+\alpha_m=d+J-1} c_{\alpha_1,\ldots,\alpha_m} \cdot n_1^{\alpha_1} \cdots n_m^{\alpha_m} \ ,$$

*is an order-$(d+J-1)$ homogeneous polynomial of $(n_1, \ldots, n_m)$ and*

$$c_{\alpha_1,\ldots,\alpha_m} = \sum_{\tau \in (\alpha_1,\ldots,\alpha_m)} |\mathbf{A}_\tau| \ . \tag{2.3}$$

The proof of Theorem 2 is relegated to the supplementary materials (Section S.3). Given a map $\tau : \{1, 2, \ldots, d+J-1\} \to \{1, \ldots, m\}$, $\mathbf{A}_\tau$ in (2.3) is a $(d+J-1) \times (d+J-1)$ matrix whose $k$th row is the same as the $k$th row of $\mathbf{A}_{\tau(k)}$, $k = 1, \ldots, d+J-1$. We take $\tau \in (\alpha_1, \ldots, \alpha_m)$ where $\alpha_i = \#\{j : \tau(j) = i\}$ for each $i = 1, \ldots, m$.

In order to obtain analytic properties of $|\mathbf{F}|$, we need some lemmas. The first of them covers Lemma 1 in Perevozskaya, Rosenberger and Haines (2003) as a special case:

**Lemma 1.** Rank$(\mathbf{A}_i)$ = Rank$(\mathbf{A}_{i3}) = J-1$. *Furthermore, $\mathbf{A}_{i3}$ is positive definite and*

$$|\mathbf{A}_{i3}| = \prod_{s=1}^{J-1} g_{is}^2 \cdot \prod_{t=1}^{J} \pi_{it}^{-1} > 0,$$

*where $g_{is} = (g^{-1})'(\theta_s - \mathbf{x}_i^T \boldsymbol{\beta}) > 0$ for $s = 1, \ldots, J-1$.*

**Example 3.** Suppose $d = 2$, $J = 3$, with link function $g$. According to Theorem 2, $|\mathbf{F}|$ is then an order-4 homogeneous polynomial of $(n_1, \ldots, n_m)$. Based on Lemma S.4 and Lemma S.5 in the supplementary materials (Section S.2), we can remove all the terms of the form $n_i^4$, $n_i^3 n_j$, or $n_i^2 n_j^2$ from $|\mathbf{F}|$. Therefore,

$$|\mathbf{F}| = \sum_{i=1}^m \sum_{j<k, j\neq i, k\neq i} c_{ijk} \cdot n_i^2 n_j n_k + \sum_{i<j<k<l} c_{ijkl} \cdot n_i n_j n_k n_l$$

for some coefficients $c_{ijk}$ and $c_{ijkl}$.

Based on Lemmas S.4 and S.5, in order to keep $c_{\alpha_1,\ldots,\alpha_m} \neq 0$, the largest possible $\alpha_i$ is $J-1$ and the fewest possible number of positive $\alpha_i$'s is $d+1$.

**Theorem 3.** $|\mathbf{F}| > 0$ *only if* $m \geq d+1$.

To determine whether $d+1$ experimental settings or support points are enough to keep the Fisher information matrix positive definite, we study the leading term of $|\mathbf{F}|$ with $\max_{1\le i\le m}\alpha_i = J-1$. For example, $a_{i_0} = J-1$ for some $1 \le i_0 \le m$. From Lemma S.5 and $\sum_{i=1}^m \alpha_i = d+J-1$, to have $c_{\alpha_1,\ldots,\alpha_m} \neq 0$, there must exist $1 \le i_1 < i_2 < \cdots < i_d \le m$ which are different from $i_0$, such that, $\alpha_{i_1} = \cdots = \alpha_{i_d} = 1$. A lemma provides an explicit formula for such a coefficient $c_{\alpha_1,\ldots,\alpha_m}$:

**Lemma 2.** *Suppose $\alpha_{i_0} = J-1$ and $\alpha_{i_1} = \cdots = \alpha_{i_d} = 1$. Then*

$$c_{\alpha_1,\ldots,\alpha_m} = \prod_{s=1}^{d} e_{i_s} \cdot |\mathbf{A}_{i_0 3}| \cdot |\mathbf{X_1}[i_0, i_1, \ldots, i_d]|^2,$$

*where $\mathbf{X_1} = (\mathbf{1}\ \mathbf{X})$ is an $m \times (d+1)$ matrix with $\mathbf{1} = (1,\ldots,1)^T$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T$, and $\mathbf{X_1}[i_0, i_1, \ldots, i_d]$ is the sub-matrix consisting of the $i_0$th, $i_1$th, $\ldots$, $i_d$th rows of $\mathbf{X_1}$.*

The proof of Lemma 2 is in the supplementary materials (Section S.3). To find D-optimal allocations, we write $|\mathbf{F}| = f(n_1, \ldots, n_m)$ for an order-$(d+J-1)$ homogeneous polynomial function $f$. The *D-optimal exact design problem* is to find an integer-valued allocation $(n_1,\ldots,n_m)$ maximizing $f(n_1,\ldots,n_m)$ subject to $n_i \in \{0,1,\ldots,n\}$, $i=1,\ldots,m$ and $n_1+\cdots+n_m = n$ with given positive integer $n$. Denote $p_i = n_i/n$, $i=1,\ldots,m$. According to Theorem 1,

$$f(n_1,\ldots,n_m) = \left|\sum_{i=1}^m n_i \mathbf{A}_i\right| = \left|n\sum_{i=1}^m p_i \mathbf{A}_i\right| = n^{d+J-1} f(p_1,\ldots,p_m). \tag{2.4}$$

Due to (2.4), Theorems 2 and 3 can be directly applied to approximate design problems too: find a real-valued allocation $(p_1,\ldots,p_m)$ maximizing $f(p_1,p_2,\ldots,p_m)$ subject to $0 \le p_i \le 1$, $i=1,\ldots,m$ and $p_1+\cdots+p_m = 1$.

According to Lemma 1, $|\mathbf{A}_{i_0 3}| > 0$. Thus $c_{\alpha_1,\ldots,\alpha_m}$ in Lemma 2 is positive as long as $\mathbf{X_1}[i_0,\ldots,i_d]$ is of full rank. Theorem 3 implies that a minimally supported design contains at least $d+1$ support points, while the following theorem states a necessary and sufficient condition for the minimum number of support points to be exactly $d+1$:

**Theorem 4.** *$f(\mathbf{p}) > 0$ for some $\mathbf{p} = (p_1,\ldots,p_m)^T$ if and only if the extended design matrix $\mathbf{X_1} = (\mathbf{1}\ \mathbf{X})$ is of full rank $d+1$.*

The minimal number of experimental settings required can thus be strictly less than the number of parameters. In the odor removal study, for example, the main-effects cumulative link model (2.1) involves four independent parameters – two $\beta$'s for the covariates ($d=2$) and two $\theta$'s for the intercepts ($J-1=2$) – while a minimally supported design could involve only three experimental settings. For multinomial responses with $J = 3$ categories, we get two degrees of freedom from each experimental setting. Here the optimal allocation of experimental units is often not uniform (see Section 4), contrary to the case of binary responses (Yang, Mandal and Majumdar (2016); Yang and Mandal (2015)).

## 3. D-Optimal Approximate Design

A (locally) D-optimal approximate design is a real-valued allocation $\mathbf{p} = (p_1, \ldots, p_m)^T$ maximizing $f(\mathbf{p}) = f(p_1, \ldots, p_m)$ with pre-specified values of parameters. The solution always exists since $f$ is continuous and the set of feasible allocations $S := \{(p_1, \ldots, p_m)^T \in \mathbb{R}^m \mid p_i \geq 0, i = 1, \ldots, m;\ \sum_{i=1}^m p_i = 1\}$ is convex and compact. A nontrivial D-optimal approximate design problem requires an assumption.

**Assumption 3.** *$m \geq d + 1$ and* $\mathrm{Rank}(\mathbf{X_1}) = d + 1$.

Assumption 3 is adopted throughout. With it, the set of *valid* allocations $S_+ := \{\mathbf{p} = (p_1, \ldots, p_m)^T \in S \mid f(\mathbf{p}) > 0\}$ is nonempty. Since $\mathbf{F} = \sum_{i=1}^m n_i \mathbf{A}_i = n \sum_{i=1}^m p_i \mathbf{A}_i$ is linear in $\mathbf{p}$ and $\phi(\cdot) = \log|\cdot|$ is concave on positive semi-definite matrices, $f(\mathbf{p}) = n^{1-d-J}|\mathbf{F}|$ is log-concave (Silvey (1980)) and thus $S_+$ is also convex.

**Theorem 5.** *A feasible allocation $\mathbf{p} = (p_1, \ldots, p_m)^T$ satisfies $f(\mathbf{p}) > 0$ if and only if* $\mathrm{Rank}(\mathbf{X_1}[\{i \mid p_i > 0\}]) = d + 1$, *where $\mathbf{X_1}[\{i \mid p_i > 0\}]$ is the sub-matrix consisting of the $\{i \mid p_i > 0\}$th rows of $\mathbf{X_1}$.*

As a direct conclusion of Theorem 5, $S_+$ contains all $\mathbf{p}$ whose coordinates are all strictly positive. A special case is the uniform allocation $\mathbf{p}_u = (1/m, \ldots, 1/m)^T$.

A necessary and sufficient condition for an approximate design to be D-optimal is of the general-equivalence-theorem type (Kiefer (1974); Pukelsheim (1993); Atkinson, Donev and Tobias (2007); Stufken and Yang (2012); Fedorov and Leonov (2014); Yang, Mandal and Majumdar (2016)), which is convenient when searching for numerical solutions. Following Yang, Mandal and Majumdar (2016), for a given $\mathbf{p} = (p_1, \ldots, p_m)^T \in S_+$ and $i \in \{1, \ldots, m\}$, we set

$$f_i(z) = f\left(\frac{1-z}{1-p_i}p_1, \ldots, \frac{1-z}{1-p_i}p_{i-1}, z, \frac{1-z}{1-p_i}p_{i+1}, \ldots, \frac{1-z}{1-p_i}p_m\right) \qquad (3.1)$$

with $0 \leq z \leq 1$. Here $f_i(z)$ is well defined as long as $p_i < 1$.

**Theorem 6.** *Suppose $\mathbf{p} = (p_1, \ldots, p_m)^T \in S_+$ and $i \in \{1, \ldots, m\}$. For $0 \leq z \leq 1$,*

$$f_i(z) = (1-z)^d \sum_{j=0}^{J-1} a_j z^j (1-z)^{J-1-j}, \tag{3.2}$$

*where* $a_0 = f_i(0)$, $(a_{J-1}, \ldots, a_1)^T = \mathbf{B}_{J-1}^{-1}\mathbf{c}$, $\mathbf{B}_{J-1} = (s^{t-1})_{s,t=1,\ldots,J-1}$, *and* $\mathbf{c} = (c_1, \ldots, c_{J-1})^T$ *with* $c_j = (j+1)^{d+J-1} j^{-d} f_i(1/(j+1)) - j^{J-1} f_i(0)$.

Following the lift-one algorithm proposed in Yang, Mandal and Majumdar (2016), we have parallel results and an algorithm for our case. For simplicity, we also call it the *lift-one algorithm*.

**Theorem 7.** *Given an allocation* $\mathbf{p} = (p_1^*, \ldots, p_m^*)^T \in S_+$, $\mathbf{p}$ *is D-optimal if and only if for each* $i = 1, \ldots, m$, $f_i(z), 0 \le z \le 1$ *attains its maximum at* $z = p_i^*$.

A lift-one algorithm

1° Start with an allocation $\mathbf{p}_0 = (p_1, \ldots, p_m)^T$ satisfying $f(\mathbf{p}_0) > 0$.

2° Set up a random order of $i$ going through $\{1, 2, \ldots, m\}$.

3° For each $i$, determine $f_i(z)$ according to Theorem 6, with $J$ determinants $f_i(0), f_i(1/2), f_i(1/3), \ldots, f_i(1/J)$ calculated according to (3.1).

4° Use the quasi-Newton method with gradient defined in (S.13) to find $z_*$ maximizing $f_i(z)$ with $0 \le z \le 1$. If $f_i(z_*) \le f_i(0)$, let $z_* = 0$. Take $\mathbf{p}_*^{(i)} = (p_1(1-z_*)/(1-p_i), \ldots, p_{i-1}(1-z_*)/(1-p_i), z_*, p_{i+1}(1-z_*)/(1-p_i), \ldots, p_m(1-z_*)/(1-p_i))^T$, so $f(\mathbf{p}_*^{(i)}) = f_i(z_*)$.

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

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

**Theorem 8.** *When the lift-one algorithm converges, the resulting* $\mathbf{p}$ *maximizes* $f(\mathbf{p})$.

**Example 4. Odor removal study** Here the response was ordinal in nature, `serious odor`, `medium odor`, and `no odor`. We fit the cumulative link model (2.1) to the data presented in Table 1. The estimated values of the model parameters are $(\hat\beta_1, \hat\beta_2, \hat\theta_1, \hat\theta_2)^T = (-2.44, 1.09, -2.67, -0.21)^T$. If a follow-up experiment is planned and the estimated parameter values are regarded as the true values, the D-optimal approximate allocation found by the lift-one algorithm is $\mathbf{p}_o = (0.4449, 0.2871,\ 0,\ 0.2680)^T$. The efficiency of the uniform $\mathbf{p}_u = (1/4, 1/4, 1/4, 1/4)^T$ is $(f(\mathbf{p}_u)/f(\mathbf{p}_o))^{1/4} = 79.7\%$, which is far from satisfactory.

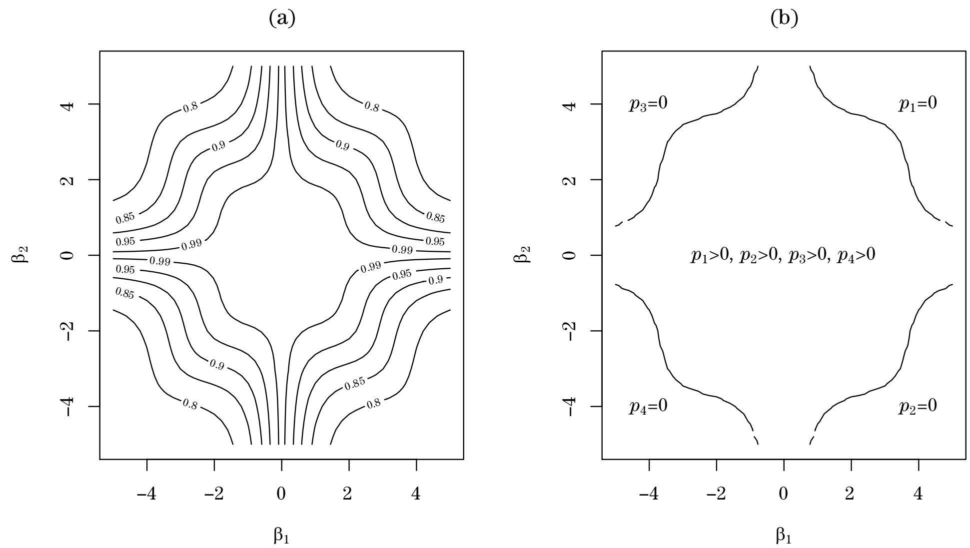


Figure 1. Wine bitterness study with assumed true parameter values $(\beta_1, \beta_2, -3.36, -0.76, 1.45, 2.99)^T$: (a) contour plot of efficiency of the original design; (b) regions for a D-optimal design to be minimally supported.

**Example 5. Wine bitterness study** (Christensen, 2015, Table 1) aggregated the wine data from Randall (1989). It contains the output of a factorial experiment with two treatment factors each at two levels (`Temperature` $x_1$: cold ($-$) or warm ($+$); `Contact` $x_2$: no ($-$) or yes ($+$)) affecting wine bitterness. The response was ordinal with five levels (from "1" being least bitter to "5" being most bitter). The original design employed a uniform allocation $\mathbf{p}_u = (1/4, 1/4, 1/4, 1/4)^T$. The estimated parameter values under the logit link are $(\hat{\beta}_1, \hat{\beta}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)^T = (1.25, 0.76, -3.36, -0.76, 1.45, 2.99)^T$. If a follow-up experiment is planned regarding the estimated values of the parameters as the true values, then the D-optimal approximate allocation found by the lift-one algorithm is $\mathbf{p}_o = (0.2694, 0.2643, 0.2333, 0.2330)^T$. The efficiency of the original design $\mathbf{p}_u$ is 99.9%. Nevertheless, the corresponding efficiency may drop to 80% if $|\beta_1|$ and $|\beta_2|$ are both larger than 3 (see Figure 1(a)). In that case, the D-optimal allocations are minimally supported, see Figure 1(b); this is discussed further in Section 5.

In the examples we have studied, the lift-one algorithm often converges within a few iterations. The Yang, Mandal and Majumdar (2016) lift-one algorithm is guaranteed to converge and can be applied if the lift-one algorithm here does not converge in a pre-specified number of iterations.

## 4. D-Optimal Exact Design

In the design literature, different discretization methods have been proposed to round an approximate design into an exact design for a given $n$, including the quota method (Kiefer (1971); Pukelsheim (1993)) and the efficient rounding procedure (Pukelsheim (1993); Pukelsheim and Rieder (1992)), which usually work well for large enough $n$ but with no guarantee for small sample size (Imhof, Lopez-Fidalgo and Wong (2001)).

In this section, we provide a direct search for D-optimal exact designs. From Theorem 5, we have the result as follows:

**Corollary 1.** $|\mathbf{F}| > 0$ *if and only if* $\mathrm{Rank}(\mathbf{X_1}[\{i \mid n_i > 0\}]) = d + 1$.

We assume $n \geq d+1$ throughout this section. To maximize $f(\mathbf{n}) = f(n_1, \ldots, n_m) = |\mathbf{F}|$, we adopt the exchange algorithm idea of Fedorov (1972). It is used here to adjust $n_i$ and $n_j$ simultaneously for randomly chosen $(i, j)$, while keeping $n_i + n_j = c$ as a constant.

We start with an $\mathbf{n} = (n_1, \ldots, n_m)^T$ satisfying $f(\mathbf{n}) > 0$. Following Yang, Mandal and Majumdar (2016), for $1 \leq i < j \leq m$, let

$$f_{ij}(z) = f\left(n_1, \ldots, n_{i-1}, z, n_{i+1}, \ldots, n_{j-1}, c - z, n_{j+1}, \ldots, n_m\right), \tag{4.1}$$

where $c = n_i + n_j$, $z = 0, 1, \ldots, c$, so $f_{ij}(n_i) = f(\mathbf{n})$. From Theorem 2, Lemmas S.4 and S.5, we have the result as follows.

**Theorem 9.** *Suppose* $\mathbf{n} = (n_1, \ldots, n_m)^T$ *satisfies* $f(\mathbf{n}) > 0$ *and* $n_i + n_j \geq J$ *for given* $1 \leq i < j \leq m$. *For* $z = 0, 1, \ldots, n_i + n_j$,

$$f_{ij}(z) = \sum_{s=0}^{J} c_s z^s, \tag{4.2}$$

*where* $c_0 = f_{ij}(0)$, *and* $c_1, \ldots, c_J$ *can be obtained using* $(c_1, \ldots, c_J)^T = \mathbf{B}_J^{-1}(d_1, \ldots, d_J)^T$ *with* $\mathbf{B}_J = (s^{t-1})_{st}$ *as a* $J \times J$ *matrix and* $d_s = (f_{ij}(s) - f_{ij}(0))/s$.

The $J \times J$ matrix $\mathbf{B}_J$ in Theorem 9 shares the same form of $\mathbf{B}_{J-1}$ in Theorem 6. According to Theorem 9, in other to maximize $f_{ij}(z)$ with $z = 0, 1, \ldots, n_i + n_j$, one can obtain the exact polynomial form of $f_{ij}(z)$ by calculating $f_{ij}(0), f_{ij}(1), \ldots, f_{ij}(J)$. There is no practical need to find out the exact form of $f_{ij}(z)$ if $n_i + n_j < J$ since one can simply calculate $f_{ij}(z)$ for each $z$. Following Yang, Mandal and Majumdar (2016), an exchange algorithm (see the supplementary materials, Section S.5) based on Theorem 9 could be used to search for a D-optimal exact allocation.

Table 2. D-optimal exact designs and the approximate design for the odor removal study.

| $n$ | $n_1$ | $n_2$ | $n_3$ | $n_4$ | $n^{-4}\lvert\mathbf{F}\rvert$ | # iterations | Time(sec.) |
|---|---|---|---|---|---|---|---|
| 3 | 1 | 1 | 0 | 1 | 0.0002911 | 1 | $<$ 0.01 |
| 10 | 4 | 3 | 0 | 3 | 0.0003133 | 3 | 0.02 |
| 40 | 18 | 11 | 0 | 11 | 0.0003177 | 3 | 0.02 |
| 100 | 44 | 29 | 0 | 27 | 0.0003180 | 4 | 0.05 |
| 1,000 | 445 | 287 | 0 | 268 | 0.0003181 | 5 | 0.39 |
| $\mathbf{p}_o$ | 0.4449 | 0.2871 | 0 | 0.2680 | 0.0003181 | 5 | 0.03 |

**Example 4 Odor removal study** (*continued*) To conduct a follow-up experiment with $n$ experimental units using the exchange algorithm, we obtain the D-optimal exact designs across different $n$'s (Table 2). As expected, the D-optimal exact allocation $(n_1, \ldots, n_4)^T$ is consistent with the D-optimal approximate allocation $\mathbf{p}_o = (p_1, \ldots, p_4)^T$ (last row of Table 2) for large $n$. The time costs in seconds (last column of Table 2) are recorded on a PC with 2GHz CPU and 8GB memory. If we rerun an experiment with $n = 40$, the D-optimal exact design is $\mathbf{n}_o = (18, 11, 0, 11)^T$, and the efficiency of the uniform design $\mathbf{n}_u = (10, 10, 10, 10)^T$ is $(f(\mathbf{n}_u)/f(\mathbf{n}_o))^{1/4} = 79.7\%$.

**Example 6. Polysilicon deposition study** Wu (2008) considered an experiment for studying the polysilicon deposition process with six 3-level factors, described in details by Phadke (1989). Due to the inconvenience of counting the number of surface defects, a major evaluating characteristic, they treated it as a 5-category ordinal variable: 1 for $0 \sim 3$ defects, 2 for $4 \sim 30$, 3 for $31 \sim 300$, 4 for $301 \sim 1{,}000$, and 5 for 1,001 and more. The original design, denoted by $\mathbf{n}_u$, includes 18 experimental settings based on an $L_{18}$ orthogonal array. To apply a cumulative link model, we represent each 3-level factor, say $A$, with levels $1, 2, 3$, by its linear component $A_1$ taking values $-1, 0, 1$ and a quadratic component $A_2$ taking values $1, -2, 1$ (Wu and Hamada (2009)). Then the fitted model with complementary log-log link chosen by both AIC and BIC criteria (see, for example, Agresti (2013)) involves four cut-points $(\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}, \hat{\alpha_4}) = (-1.59, -0.58, 0.41, 1.22)$, and twelve other coefficients $(\hat{\beta_{11}}, \hat{\beta_{12}}, \hat{\beta_{21}}, \hat{\beta_{22}}, \ldots, \hat{\beta_{62}}) = (1.45, -0.22, 1.35, 0.02, -0.12, -0.34, 0.19, 0.00, 0.22,\ 0.08,\ 0.05,\ 0.17)$. When the true parameter values were assumed to be the estimated ones, we use the exchange algorithm to find a D-optimal 18-run design, denoted by $\mathbf{n}_o$ (see the supplementary materials, Section S.6, for a list of the 18 experimental settings). Compared with $\mathbf{n}_o$, the efficiency of the original design $\mathbf{n}_u$ is $(f(\mathbf{n}_u)/f(\mathbf{n}_o))^{1/16} = 73.1\%$. In order to check the efficiency of a rounded design, we use the lift-one algorithm to find that the D-optimal approximate design contains 100 positive $p_i$'s out of the 729 distinct experimental settings. In this case, both the quota method and the efficient rounding procedure end with the same rounded design $\mathbf{n}_r$ (see Section S.6). Its efficiency is $(f(\mathbf{n}_r)/f(\mathbf{n}_o))^{1/16} = 86.1\%$.

## 5. Minimally Supported Design

It is of practical significance to have an experiment run with the minimal number of different settings. For example, the 18 experimental settings in the polysilicon deposition study (Example 6) had to be run in a sequential way and only two settings were arranged on each day (Phadke (1989)). Less experimental settings often indicate less time and less cost. Another practical application of a minimally supported design is that an optimal allocation restricted to those support points can be obtained more easily or even analytically.

According to Theorem 3, a minimally supported design contains at least $d+1$ support points. On the other hand, according to Theorem 5 and Corollary 1, a minimally supported design could contain exactly $d+1$ support points if the extended design matrix $\mathbf{X_1} = (\mathbf{1}\ \mathbf{X})$ is of full rank.

**Example 7.** Let $J = 2$ with a binomial response. There are $d+1$ parameters, $\theta_1, \beta_1, \ldots, \beta_d$. For a general link function $g$ satisfying Assumptions 1 and 2, $g_{i0} = g_{i2} = 0$, $g_{i1} = (g^{-1})'(\theta_1 - \mathbf{x}_i^T\boldsymbol{\beta}) > 0$, $e_i = u_{i1} = c_{i1} = g_{i1}^2/[\pi_{i1}(1-\pi_{i1})]$, $i = 1, \ldots, m$. Then $\mathbf{A}_{i3}$ in Theorem 1 contains only the entry $u_{i1}$, and thus $|\mathbf{A}_{i3}| = u_{i1}$, or simply $e_i$ (Lemma 1 still holds). Assume further that the $m \times d$ design matrix $\mathbf{X}$ satisfies Assumption 3. According to Theorem 2, Lemmas S.4, S.5, and 2, given $\mathbf{p} = (p_1, \ldots, p_m)^T$,

$$f(\mathbf{p}) = n^{-(d+1)}|\mathbf{F}| = \sum_{1 \le i_0 < i_1 < \cdots < i_d \le m} |\mathbf{X_1}[i_0, i_1, \ldots, i_d]|^2 p_{i_0} e_{i_0} p_{i_1} e_{i_1} \cdots p_{i_d} e_{i_d}. \tag{5.1}$$

Here (5.1) is essentially the same as Lemma 3.1 in Yang and Mandal (2015). Then a minimally supported design can contain $d+1$ support points and a D-optimal one keeps equal weight $1/(d+1)$ on all support points (Yang and Mandal (2015, Thm. 3.2)).

For univariate responses (including binomial ones) under a generalized linear model, a minimally supported design must keep equal weights on all its support points in order to keep D-optimality (Yang and Mandal (2015)). However, for multinomial responses with $J \ge 3$, this is usually not the case. In this section, we use one-predictor ($d = 1$) and two-predictor ($d = 2$) cases for illustration.

In order to check if a minimally supported design is D-optimal, we need a Karush-Kuhn-Tucker-type condition. Since $f(\mathbf{p})$ is log-concave, the Karush-Kuhn-Tucker conditions (Karush (1939); Kuhn and Tucker (1951)) are also sufficient.

**Theorem 10.** *An allocation $\mathbf{p} = (p_1^*, \ldots, p_m^*)^T$ satisfying $f(\mathbf{p}) > 0$ is D-optimal if and only if there exists a $\lambda \in \mathbb{R}$ such that $\partial f(\mathbf{p})/\partial p_i = \lambda$ if $p_i^* > 0$ or $\leq \lambda$ if $p_i^* = 0$, $i = 1, \ldots, m$.*

### 5.1. Minimally supported designs with one predictor

We start with $d = 1$ and $J \geq 3$. The corresponding parameters here are $\beta_1$ and $\theta_1, \ldots, \theta_{J-1}$. Consider designs supported on two points ($m = 2$, minimally supported), and invoke Theorem 2, Lemmas S.4 and S.5.

**Theorem 11.** *If $d = 1$, $J \geq 3$, and $m = 2$, the objective function is*

$$f(p_1, p_2) = n^{-2}|\mathbf{F}| = \sum_{s=1}^{J-1} c_s p_1^{J-s} p_2^s, \tag{5.2}$$

*where $(c_1, \ldots, c_{J-1})^T = \mathbf{B}_{J-1}^{-1}(d_1, \ldots, d_{J-1})^T$, with $\mathbf{B}_{J-1} = (s^{t-1})_{st}$ as a $(J-1) \times (J-1)$ matrix and $d_s = f(1/(s+1), s/(s+1)) \cdot (s+1)^J/s$.*

Actually, according to Lemma 2, $c_1 = e_2 \prod_{s=1}^{J-1} g_{1s}^2 \cdot \prod_{t=1}^{J} \pi_{1t}^{-1}(x_1 - x_2)^2$, $c_{J-1} = e_1 \prod_{s=1}^{J-1} g_{2s}^2 \cdot \prod_{t=1}^{J} \pi_{2t}^{-1}(x_1 - x_2)^2$, where $x_1, x_2$ are the predictor levels. Theorem 11 provides a way to find the exact form of $f(p_1, p_2)$ after calculating $|\mathbf{F}|$ for $J - 1$ different allocations. Then the D-optimal problem is to maximize an order-$J$ polynomial $f(z, 1-z)$ for $z \in [0, 1]$. As a special case, the D-optimal allocation of $J = 3$ can be solved explicitly as follows:

**Corollary 2.** *If $d = 1$, $J = 3$, and $m = 2$, the objective function is*

$$f(p_1, p_2) = p_1 p_2 (c_1 p_1 + c_2 p_2), \tag{5.3}$$

*where $c_1 = e_2 g_{11}^2 g_{12}^2 (\pi_{11}\pi_{12}\pi_{13})^{-1}(x_1 - x_2)^2 > 0$, $c_2 = e_1 g_{21}^2 g_{22}^2 (\pi_{21}\pi_{22}\pi_{23})^{-1}$ $(x_1 - x_2)^2 > 0$, and $x_1, x_2$ are the two levels of the predictor. The D-optimal design $\mathbf{p} = (p_1^*, p_2^*)$ is*

$$p_1^* = \frac{c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}}{2c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}}\ , \quad p_2^* = \frac{c_1}{2c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}}. \tag{5.4}$$

*Furthermore, $p_1^* = p_2^* = 1/2$ if and only if $c_1 = c_2$.*

Under the setup of Corollary 2, $p_1^* = p_2^* = 1/2$ if $\beta_1 = 0$. In general $p_1^* \neq p_2^*$, and $p_1^* > p_2^*$ if and only if $c_1 > c_2$. The following result provides conditions for

D-optimality of such a minimally supported design. Its proof is relegated to the supplementary materials (Section S.3).

**Corollary 3.** *Suppose $d = 1$, $J = 3$, $m \geq 3$, and let $x_1, \ldots, x_m$ be the $m$ distinct levels of the predictor. A minimally supported design $\mathbf{p} = (p_1^*, p_2^*, 0, \ldots, 0)^T$ is D-optimal if and only if*

(1) *$p_1^*, p_2^*$ are defined as in* (5.4),

(2) *$s_{i3}(p_1^*)^2 + (s_{i5} - 2c_1)p_1^* p_2^* + (s_{i4} - c_2)(p_2^*)^2 \leq 0$, $i = 3, \ldots, m$,*

*where $c_1, c_2$ are as in Corollary 2, $s_{i3} = e_i g_{11}^2 g_{12}^2 (\pi_{11}\pi_{12}\pi_{13})^{-1}(x_1 - x_i)^2 > 0$, $s_{i4} = e_i g_{21}^2 g_{22}^2 (\pi_{21}\pi_{22}\pi_{23})^{-1}(x_2 - x_i)^2 > 0$, $s_{i5} = e_1(u_{22}u_{i1} + u_{21}u_{i2} - 2b_{22}b_{i2})(x_1 - x_2)(x_1 - x_i) + e_2(u_{12}u_{i1} + u_{11}u_{i2} - 2b_{12}b_{i2})(x_2 - x_1)(x_2 - x_i) + e_i(u_{12}u_{21} + u_{11}u_{22} - 2b_{12}b_{22})(x_i - x_1)(x_i - x_2)$.*

**Example 8.** Consider $d = 1$, $J = 3$, and $m = 3$ with factor levels $\{-1, 0, 1\}$. Under the logit link $g$, the parameters $\beta, \theta_1, \theta_2$ satisfy $g(\gamma_{1j}) = \theta_j + \beta$, $g(\gamma_{2j}) = \theta_j$, $g(\gamma_{3j}) = \theta_j - \beta$, $j = 1, 2$. We investigate when a D-optimal design is minimally supported. According to Theorem 11, a D-optimal deign satisfies $p_1 = p_3 = 1/2$ if $\beta = 0$. Figure 2 shows cases with more general parameter values. In Figure 2(a), four regions in $(\theta_1, \theta_2)$-plane are occupied by minimally supported designs ($\theta_1 < \theta_2$ is required). For example, regions labeled with $p_2 = 0$ indicates a minimally supported design satisfying $p_2 = 0$ is D-optimal given such a triple $(\theta_1, \theta_2, \beta = -2)$. From Figure 2(b), a design supported on $\{-1, 1\}$ (that is, $p_2 = 0$) is D-optimal if $\beta$ is not far from 0.

**Example 9. Toxicity study** (Agresti, 2013, Table 8.7) reported data from a developmental toxicity study with one factor (concentration of diEGdiME at five levels: 0, 62.5, 125, 250, 500 mg/kg per day) and a 3-category ordinal response (status of mouse fetus: `nonlive`, `malformation`, or `normal`). In this case, $d = 1$, $J = 3$, and $m = 5$. We fit a cumulative link model with `cauchit` link chosen by both AIC and BIC criteria. The estimated parameter values are $(\hat{\beta}_1, \hat{\theta}_1, \hat{\theta}_2)^T = (-0.0176, -8.80, -5.34)$. If $(\hat{\beta}_1, \hat{\theta}_1, \hat{\theta}_2)^T$ is regarded as the true parameter value, then the D-optimal approximate allocation found by the lift-one algorithm is $\mathbf{p}_o = (0, 0, 0, 0.4285, 0.5715)^T$, which is minimally supported. Alternatively, for each pair of indices $(i, j)$, $1 \leq i < j \leq 5$, we obtain the best design $(p_i^*, p_j^*)$ supported only on $x_i, x_j$ according to Corollary 2, then check whether $(p_i^*, p_j^*)$ is D-optimal using Corollary 3. Here $\mathbf{p}_o$ is the only minimally supported design that is also D-optimal. With respect to $\mathbf{p}_o$, the efficiency of the original design (roughly a uniform one) is 52.6%.

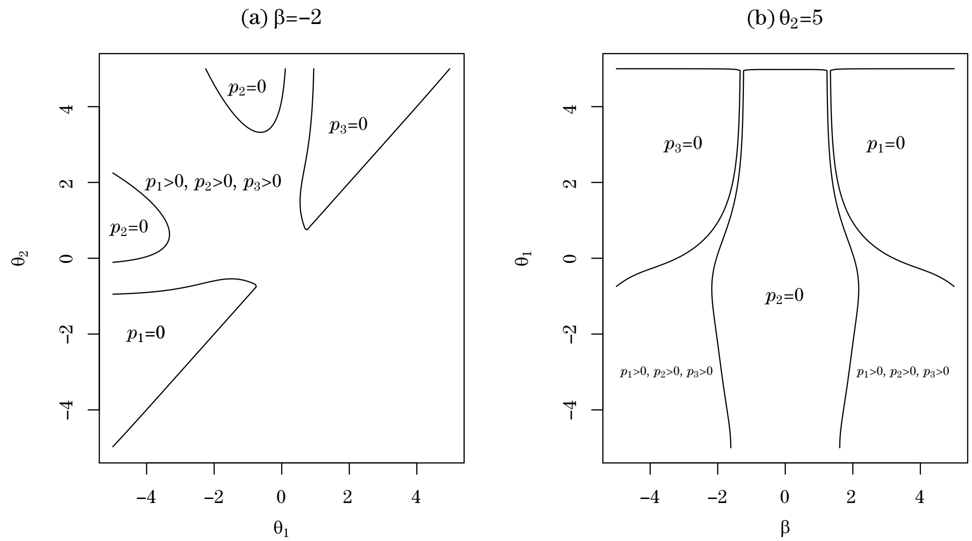


Figure 2. Regions for a two-point design to be D-optimal with $d = 1$, $J = 3$, $x \in \{-1, 0, 1\}$, and logit link (note that $\theta_1 < \theta_2$ is required).

## 5.2. Minimally supported designs with two predictors

In this section, we consider experiments with two predictors ($d = 2$) and a three-category response ($J = 3$). The parameters are $\beta_1, \beta_2, \theta_1, \theta_2$. For cases with $J \geq 4$, similar conclusions could be obtained, but with messier notation.

According to Theorem 3, a minimally supported design needs three support points, for example, $(x_{i1}, x_{i2})$, $i = 1, 2, 3$. Under Assumption 3, the $3 \times 3$ matrix $\mathbf{X_1} = (\mathbf{1}\ \mathbf{X})$ is of full rank. Following Theorem 2, Lemmas S.4, S.5, and 2, the objective function with $(d, J, m) = (2, 3, 3)$ is

$$f(p_1, p_2, p_3) = |\mathbf{X_1}|^2 e_1 e_2 e_3 \cdot p_1 p_2 p_3 (w_1 p_1 + w_2 p_2 + w_3 p_3), \tag{5.5}$$

where $w_i = e_i^{-1} g_{i1}^2 g_{i2}^2 (\pi_{i1}\pi_{i2}\pi_{i3})^{-1} > 0$. Since $f(p_1,\, p_2,\, p_3) = 0$ if $p_1 p_2 p_3 = 0$, we need only consider $\mathbf{p} = (p_1, p_2, p_3)^T$ satisfying $0 < p_1, p_2, p_3 < 1$.

According to Theorem 10, $\mathbf{p}$ maximizes $f(p_1, p_2, p_3)$ only if

$$\frac{\partial f}{\partial p_1} = \frac{\partial f}{\partial p_2} = \frac{\partial f}{\partial p_3}. \tag{5.6}$$

Following Tong, Volkmer and Yang (2014), we obtain its analytic solution:

**Theorem 12.** *Without loss of generality, $w_1 \geq w_2 \geq w_3 > 0$. The allocation $\mathbf{p} = (p_1^*, p_2^*, p_3^*)^T$ maximizing $f(p_1, p_2, p_3)$ in (5.5) exists and is unique. It satisfies $0 < p_3^* \leq p_2^* \leq p_1^* < 1$ and can be obtained analytically as follows:*

(i) *If $w_1 \geq w_2 = w_3$, then $p_1^* = \Delta_1/(4w_1 + \Delta_1)$, $p_2^* = p_3^* = 2w_1/(4w_1 + \Delta_1)$,* *where* $\Delta_1 = 2w_1 - 3w_2 + \sqrt{4w_1^2 - 4w_1w_2 + 9w_2^2}$*. A special case is* $p_1^* = p_2^* = p_3^* = 1/3$ *if* $w_1 = w_2 = w_3$.

(ii) *If* $w_1 = w_2 > w_3$*, then* $p_1^* = p_2^* = \Delta_2/[2(\Delta_2 + 2w_1)]$*,* $p_3^* = 2w_1/(\Delta_2 + 2w_1)$*, where* $\Delta_2 = 3w_1 - 2w_3 + \sqrt{9w_1^2 - 4w_1w_3 + 4w_3^2}$.

(iii) *If* $w_1 > w_2 > w_3$*, then* $p_1^* = y_1/(y_1 + y_2 + 1)$*,* $p_2^* = y_2/(y_1 + y_2 + 1)$*,* $p_3^* = 1/(y_1 + y_2 + 1)$*, where*

$$y_1 = -\frac{b_2}{3} - \frac{2^{1/3}(3b_1 - b_2^2)}{3A^{1/3}} + \frac{A^{1/3}}{3 \times 2^{1/3}}, \quad y_2 = \frac{(w_1 - w_3)y_1}{(w_2 - w_3) + (w_1 - w_2)y_1}$$

*with* $A = -27b_0 + 9b_1b_2 - 2b_2^3 + 3^{3/2}(27b_0^2 + 4b_1^3 - 18b_0b_1b_2 - b_1^2b_2^2 + 4b_0b_2^3)^{1/2}$*,* $b_i = c_i/c_3$*,* $i = 0, 1, 2$*, and* $c_0 = w_3(w_2 - w_3) > 0$*,* $c_1 = 3w_1w_2 - w_1w_3 - 4w_2w_3 + 2w_3^2 > 0$*,* $c_2 = 2w_1^2 - 4w_1w_2 - w_1w_3 + 3w_2w_3$*,* $c_3 = w_1(w_2 - w_1) < 0$.

The proof of Theorem 12 is relegated to the supplementary materials (Section S.3).

**Corollary 4.** *Suppose* $d = 2$*,* $J = 3$*, and* $m = 3$*. Then* $\mathbf{p} = (1/3, 1/3, 1/3)^T$ *is D-optimal if and only if* $w_1 = w_2 = w_3$*, where* $w_1, w_2, w_3$ *are defined as in* (5.5).

**Example 10.** Consider a $2^2$ factorial design problem with a three-category response and four design points $(1, 1)$, $(1, -1)$, $(-1, 1), (-1, -1)$, denoted by $(x_{i1}, x_{i2}), i = 1, 2, 3, 4$. Take $w_i = e_i^{-1} g_{i1}^2 g_{i2}^2 (\pi_{i1}\pi_{i2}\pi_{i3})^{-1}$, $i = 1, 2, 3, 4$. There are five special cases: *(i)* if $\beta_1 = \beta_2 = 0$, then $w_1 = w_2 = w_3 = w_4$; *(ii)* if $\beta_1 = 0, \beta_2 \neq 0$, then $w_1 = w_3$, $w_2 = w_4$, but $w_1 \neq w_2$; *(iii)* if $\beta_1 \neq 0, \beta_2 = 0$, then $w_1 = w_2$, $w_3 = w_4$, but $w_1 \neq w_3$; *(iv)* if $\beta_1 = \beta_2 \neq 0$, then $w_2 = w_3$, but $w_1, w_2, w_4$ are distinct; *(v)* if $\beta_1 = -\beta_2 \neq 0$, then $w_1 = w_4$, but $w_1, w_2, w_3$ are distinct.

Theorem 12 provides analytic forms of minimally supported designs with $d = 2$ and $J = 3$.

**Corollary 5.** *Suppose* $d = 2$*,* $J = 3$*, and* $m \geq 4$*. Let* $(x_{i1}, x_{i2})$*,* $i = 1, \ldots, m$ *be m distinct level combinations of the two predictors. With* $\mathbf{X_1} = (\mathbf{1}\ \mathbf{X})$ *an* $m \times 3$ *matrix, a minimally supported design* $\mathbf{p} = (p_1^*, p_2^*, p_3^*, 0, \ldots, 0)^T$ *is D-optimal if and only if* $p_1^*, p_2^*, p_3^*$ *are obtained according to Theorem* 12*, and*

$$\begin{aligned}
&|\mathbf{X_1}[1,2,i]|^2 e_1e_2e_i p_1^*p_2^*(w_1p_1^* + w_2p_2^*) + |\mathbf{X_1}[1,3,i]|^2 e_1e_3e_i p_1^*p_3^*(w_1p_1^* + w_3p_3^*)\\
&+ |\mathbf{X_1}[2,3,i]|^2 e_2e_3e_i p_2^*p_3^*(w_2p_2^* + w_3p_3^*) + D_i p_1^*p_2^*p_3^*\\
&\leq |\mathbf{X_1}[1,2,3]|^2 e_1e_2e_3 p_2^*p_3^*(2w_1p_1^* + w_2p_2^* + w_3p_3^*), \qquad \text{for } i = 4, \ldots, m,
\end{aligned}$$

*where* $e_j = u_{j1} + u_{j2} - 2b_{j2}$*,* $w_j = e_j^{-1} g_{j1}^2 g_{j2}^2 (\pi_{j1}\pi_{j2}\pi_{j3})^{-1}$*,* $j = 1, \ldots, m$*,* $D_i =$

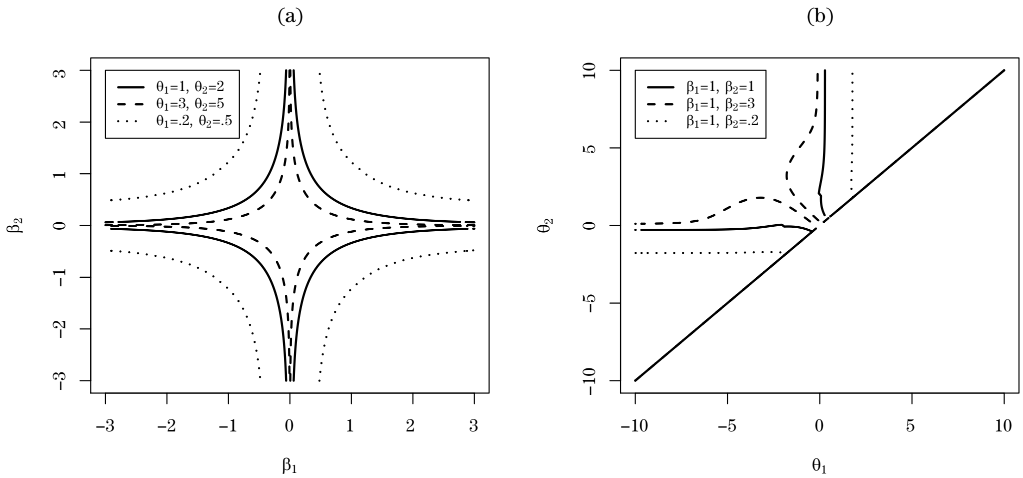


Figure 3. Boundary lines for a three-point design to be D-optimal with logit link: Region of $(\beta_1, \beta_2)$ for given $(\theta_1, \theta_2)$ is outside the boundary lines in Panel (a); Region of $(\theta_1, \theta_2)$ (with $\theta_1 < \theta_2$) for given $(\beta_1, \beta_2)$ is between the boundary lines and $\theta_1 = \theta_2$ in Panel (b).

$\sum_{\{j,k,s,t\}\in E_i} e_j e_k (u_{s1}u_{t2} + u_{s2}u_{t1} - 2b_{s2}b_{t2}) \cdot |\mathbf{X_1}[j,k,s]| \cdot |\mathbf{X_1}[j,k,t]|$ *with the sum over* $E_i = \{(1,2,3,i), (1,3,2,i), (1,i,2,3),\ (2,3,1,i),\ (2,i,1,3), (3,i,1,2)\}$.

**Example 11.** Consider experiments with $d = 2$, $J = 3$, $m = 4$, and design points $(1,\ 1)$, $(1,\ -1)$, $(-1,\ 1)$, $(-1, -1)$. Figure 3 provides the boundary lines of regions of parameters $(\beta_1, \beta_2, \theta_1, \theta_2)$ for which the best three-point design is D-optimal. In particular, Figure 3(a) shows the region of $(\beta_1, \beta_2)$ for given $\theta_1, \theta_2$. It clearly indicates that the best three-point design tends to be D-optimal when the absolute values of $\beta_1, \beta_2$ are large. The region tends to be larger as the absolute values of $\theta_1, \theta_2$ increase. On the other hand, Figure 3(b) displays the region of $(\theta_1, \theta_2)$ for given $\beta_1, \beta_2$. The symmetry of the boundary lines about $\theta_1 + \theta_2 = 0$ is due to the logit link which is symmetric about 0. An interesting conclusion based on Corollary 5 is that in this case a three-point design can never be D-optimal if $\beta_1 = 0$ or $\beta_2 = 0$.

**Remark 1.** Extra degrees of freedom play an important role against the uniformity of D-optimal allocation in a minimally supported design. For multinomial-type responses with $J$ categories, the total degrees of freedom from $m$ distinct experimental settings is $m(J-1)$, while a cumulative link model contains $d + J - 1$ parameters. For a minimally supported design, $m = d + 1$ and $m(J-1) = d + J - 1$ if and only if $J = 2$ (see Example 7). Then the objective function $f(\mathbf{p}) \propto p_{i_0} p_{i_1} \cdots p_{i_d}$ and the D-optimal allocation is $p_{i_0} = p_{i_1} =$

$\cdots = p_{i_d} = 1/(d+1)$. However, if $J \geq 3$, the degrees of freedom is strictly larger than the number of parameters and there are "extra" degrees of freedom. In this case, distinct experimental settings may play different roles in estimating the parameters values. For example, if $d = 1, J = 3, m = 2$, the objective function $f(\mathbf{p}) = p_1 p_2 (c_1 p_1 + c_2 p_2)$ according to Corollary 2; if $d = 2, J = 3, m = 3$, $f(\mathbf{p}) \propto p_1 p_2 p_3 (w_1 p_1 + w_2 p_2 + w_3 p_3)$ according to equation (5.5). The D-optimality of a uniform allocation requires $c_1 = c_2$ or $w_1 = w_2 = w_3$, which is not true in general.

## 6. EW D-Optimal Design

The previous sections mainly focus on locally D-optimal designs which require assumed parameter values, $(\beta_1, \ldots, \beta_d, \theta_1, \ldots, \theta_{J-1})$. For many applications, the experimenter may have little information about the values of parameters. Then Bayes D-optimality (Chaloner and Verdinelli (1995)) which maximizes $E(\log |\mathbf{F}|)$ given a prior distribution on parameters provides a reasonable solution. An alternative is EW D-optimality (Yang, Mandal and Majumdar (2016); Atkinson, Donev and Tobias (2007)) which essentially maximizes $\log |E(\mathbf{F})|$. According to Yang, Mandal and Majumdar (2016)'s simulation study across different models and choices of priors, EW D-optimal designs are much easier to calculate and still highly efficient compared with Bayes designs.

Based on Theorem 1, an EW D-optimal design that maximizes $|E(\mathbf{F})|$ can be viewed as a locally D-optimal design with $e_i, c_{it}, u_{it}$ and $b_{it}$ replaced by their expectations. After the replacement, Lemma S.2 still holds. Therefore, almost all results in the previous sections can be applied directly to EW D-optimal designs. The only exception is Lemma 1 which provides the formula for $|\mathbf{A}_{i3}|$ in terms of $g_{ij}$ and $\pi_{ij}$. In order to find EW D-optimal designs, $|\mathbf{A}_{i3}|$ needs to be calculated in terms of $u_{it}$ and $b_{it}$. For example, $|\mathbf{A}_{i3}| = u_{i1}$ if $J = 2$, $|\mathbf{A}_{i3}| = u_{i1}u_{i2} - b_{i2}^2$ if $J = 3$, and $|\mathbf{A}_{i3}| = u_{i1}u_{i2}u_{i3} - u_{i1}b_{i3}^2 - u_{i3}b_{i2}^2$ if $J = 4$. Then the formulas of $|\mathbf{A}_{i3}|$ in Lemma 2, $c_1, c_2$ in Corollary 2, $s_{i3}, s_{i4}, s_{i5}$ in Corollary 3, $w_i$ in (5.5), and $w_j$ in Corollary 5 need to be written in terms of $u_{it}$ and $b_{it}$.

According to Lemma S.2, we only need to calculate $E(u_{it}), i = 1, \ldots, m; t = 1, \ldots, J-1$ and $E(b_{it}), i = 1, \ldots, m; t = 2, \ldots, J-1$ (if $J \geq 3$). Then $E(c_{it}) = E(u_{it}) - E(b_{it}) - E(b_{i,t+1})$ and $E(e_i) = \sum_{t=1}^{J-1} E(c_{it})$. After that, we can use the lift-one algorithm in Section 3 or the exchange algorithm in Section 4 to find EW D-optimal designs.

**Example 4: Odor Removal Study** (*continued*) Instead of assuming the parameter values $(\beta_1, \beta_2, \theta_1, \theta_2) = (-2.44, 1.09, -2.67, -0.21)$, consider true val-

Table 3. Summary of efficiency in odor removal study.

| Design | Min. | 1st Quartile | Median | Mean | 3rd Quartile | Max. |
|---|---|---|---|---|---|---|
| Bayes $\mathbf{p}_b$ | 0.8464 | 0.9813 | 0.9915 | 0.9839 | 0.9964 | 1.0000 |
| EW $\mathbf{p}_e$ | 0.8465 | 0.9802 | 0.9917 | 0.9838 | 0.9967 | 1.0000 |
| Uniform $\mathbf{p}_u$ | 0.7423 | 0.8105 | 0.8622 | 0.8674 | 0.9249 | 0.9950 |

ues of parameters that satisfy $\beta_1 \in [-3,-1]$, $\beta_2 \in [0,2]$, $\theta_1 \in [-4,-2]$, and $\theta_2 \in [-1,1]$. We assume that the four parameters are independently and uniformly distributed within their intervals. We use `R` function `constrOptim` to maximize $\phi(\mathbf{p}) = E(\log|\mathbf{F}|)$ and find the Bayes D-optimal allocation $\mathbf{p}_b = (0.3879, 0.3264, 0.0000, 0.2857)^T$. The procedure costs 313 seconds computational time using a PC with 2GHz CPU and 8GB memory. In order to get the EW D-optimal design, we only need 5.43 seconds in total to calculate $E(u_{it})$, $E(b_{it})$, and find $\mathbf{p}_e = (0.3935, 0.3259,\ 0,\ 0.2806)^T$ using the lift-one algorithm. Even in terms of Bayes Optimality (Chaloner and Larntz (1989); Song and Wong (1998); Abebe et al. (2014)), the relative efficiency of $\mathbf{p}_e$ with respect to $\mathbf{p}_b$ is $\exp\{(\phi(\mathbf{p}_e) - \phi(\mathbf{p}_b))/4\} \times 100\% = 99.99\%$, while the relative efficiency of the uniform allocation $\mathbf{p}_u = (0.25, 0.25, 0.25, 0.25)^T$ is 87.67%.

In order to check *robustness* towards misspecified parameter values, we let $\boldsymbol{\theta} = (\beta_1, \beta_2, \theta_1, \theta_2)^T$ run through all 0.1-grid points in $[-3,-1]\times[0,2]\times[-4,-2]\times[-1,1]$. For each $\boldsymbol{\theta}$, we use the lift-one algorithm to find the D-optimal allocation $\mathbf{p}_{\boldsymbol{\theta}}$ and the corresponding determinant $f(\mathbf{p}_{\boldsymbol{\theta}}) = |\mathbf{F}(\mathbf{p}_{\boldsymbol{\theta}})|$, and then calculate the efficiency $(f(\mathbf{p})/f(\mathbf{p}_{\boldsymbol{\theta}}))^{1/4}$ for $\mathbf{p} = \mathbf{p}_b, \mathbf{p}_e$, and $\mathbf{p}_u$, respectively. Table 3 shows the summary statistics of the efficiencies. It implies that $\mathbf{p}_b$ and $\mathbf{p}_e$ are comparable and both of them are much better than $\mathbf{p}_u$ in terms of robustness.

## 7. Discussion

In this paper, we use real experiments to illustrate how much improvements the experimenter could make. Compared with our D-optimal designs, the efficiencies of the original designs are often far from satisfactory: 79.7% in Example 4, 73.1% in Example 6, and 52.6% in Example 9. More interestingly, our D-optimal designs recommended for Example 4 and Example 9 are both minimally supported. We have two surprising findings that are different from the cases under univariate generalized linear models (Yang and Mandal, 2015): (1) the minimum number of experimental settings can be strictly less than the number of parameters, and (2) the allocation of experimental units on the support points of a minimally supported design is usually not uniform.

Cumulative link models are widely used for modeling ordinal data. Nevertheless, there are other models used for multinomial-type responses, including baseline-category logit model for nominal response, adjacent-categories logit model for ordinal data, and continuation-ratio logit model for hierarchical response (see Liu and Agresti (2005), Agresti (2013) for a review). The methods developed in this paper could be extended for those models as well. For further extensions, our approaches could be used for planning experiments with more than one categorical response. For example, both the paper feeder experiment and the PCB experiment analyzed by Joseph and Wu (2004) involved multiple binomial responses.

## Supplementary Materials

The proofs of Theorems 1, 2, 4, 5, and 12, Lemma 2, and Corollaries 3 and 5 are available at `http://www3.stat.sinica.edu.tw/statistica/`. There are also tabularized formulas for commonly used link functions, additional lemmas for Section 2 and Section 5.2, maximization of $f_i(z)$ in Section 3, exchange algorithm for D-optimal exact allocation in Section 4, and more results for Example 4.6.

## Acknowledgment

We thank Dr. Suraj Sharma for providing the details of the odor removal study, and Dr. John Stufken for valuable suggestions on an early version of this paper. We also thank an associate editor and the reviewers for comments and suggestions that substantially improved the quality of the manuscript. This research is in part supported by the LAS Award for Faculty of Science at UIC.

# References

Abebe, H. T., Tan, F. E., Van Breukelen, G. J., Serroyen, J. and Berger, M. P. (2014). On the choice of a prior for Bayesian D-optimal designs for the logistic regression model with a single predictor. *Communications in Statistics - Simulation and Computation* **43**, 1,811-1,824.

Agresti, A. (2013). *Categorical Data Analysis*, Third Edition. Wiley, New Jersey.

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

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference* **21**, 191-208.

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.

Christensen, R. H. B. (2015). Analysis of ordinal data with cumulative link models – estimation with the R-package ordinal. Available via `http://cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf`

Dobson, A. J. and Barnett, A. (2008). *An Introduction to Generalized Linear Models*, Third Edition. Chapman & Hall/CRC, London.

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

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

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

Ford, I., Torsney, B. and Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society, Series B* **54**, 569-583.

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

Imhof, L., Lopez-Fidalgo, J. and Wong, W.K. (2001). Efficiencies of rounded optimal approximate designs for small samples. *Statistica Neerlandica* **55**, 301-318.

Joseph, V. R. and Wu, C. F. J. (2004). Failure amplification method: an information maximization approach to categorical response optimization (with discussions). *Technometrics* **46**, 1-31.

Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints, M.Sc. Dissertation, Department of Mathematics, University of Chicago.

Kiefer, J. (1971). The role of symmetry and approximation in exact design optimality. In *Statistical Decision Theory and Related Topics*, S.S. Gupta and J. Yackel (eds), 109-118, Academic Press, New York.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics* **2**, 849-879.

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

Krause, M. S., Madden, L. V. and Hoitink, H. A. J. (2001). Effect of potting mix microbial carrying capacity on biological control of Rhizoctonia damping-off of radish and Rhizoctonia crown and root rot of Poinsettia. *Phytopathology* **91**, 1,116-1,123.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. *Proceedings of 2nd Berkeley Symposium*, Berkeley: University of California Press, 481-492.

Liu, I. and Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test* **14**, 1-73.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B* **42**, 109-142.

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

Omer, M. A., Johnson, D. A. and Rowe, R. C. (2000). Recovery of *Verticillium dahliae* from North American certified seed potatoes and characterization of strains by vegetative compatibility and aggressiveness. *American Journal of Potato Research* **77**, 325-331.

Osterstock, J. B., MacDonald, J. C., Boggess, M. M. and Brown, M. S. (2010). Analysis of ordinal outcomes from carcass data in beef cattle research. *Journal of Animal Science* **88**, 3384-3389.

Perevozskaya, I., Rosenberger, W. F. and Haines, L. M. (2003). Optimal design for the proportional odds model. *The Canadian Journal of Statistics* **31**, 225-235.

Phadke, M. S. (1989). *Quality Engineering using Robust Design*. Prentice-Hall, Englewood Cliffs.

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

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

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79**, 763-770.

Randall, J. (1989). The analysis of sensory data by generalised linear model. *Biometrical Journal* **31**, 781-793.

Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

Song, D. and Wong, W. K. (1998). Optimal two-point designs for the Michaelis-Menten model with heteroscedastic errors. *Communications in Statistics - Theory and Methods* **27**, 1,503-1,516.

Stufken, J. and Yang, M. (2012). Optimal designs for generalized linear models. In: *Design and Analysis of Experiments*, Volume 3: Special Designs and Applications, K. Hinkelmann (ed.). Wiley, New York.

Tong, L., Volkmer, H. W. and Yang, J. (2014). Analytic solutions for D-optimal factorial designs under generalized linear models. *Electronic Journal of Statistics* **8**, 1,322-1,344.

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.

Wu, F-C. (2008). Simultaneous optimization of robust design with quantitative and ordinal data. *International Journal of Industrial Engineering: Theory, Applications and Practice* **15**, 231-238.

Wu, C. F. J. and Hamada, M. (2009). *Experiments: Planning, Analysis, and Optimization*, Second Edition. Wiley, New York.

Yang, J. and Mandal, A. (2015). D-optimal factorial designs under generalized linear models. *Communications in Statistics - Simulation and Computation* **44**, 2,264-2,277.

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

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

Dept. of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 S Morgan Street (M/C 249), Chicago, IL 60607, USA

E-mail: jyang06@math.uic.edu

Advocate Health Care, 3075 Highland Parkway, Suite 600, Downers Grove, IL 60515, USA

E-mail: liping.tong@advocatehealth.com

Department of Statistics, University of Georgia, 101 Cedar Street, Athens, GA 30602 - 7952, USA

E-mail: amandal@stat.uga.edu

(Received April 2016; accepted October 2016)