

OPTIMAL DESIGNS FOR GENERALIZED LINEAR MODELS WITH MULTIPLE DESIGN VARIABLES

Min Yang, Bin Zhang and Shuguang Huang

*University of Missouri, University of Alabama-Birmingham and
Precision Therapeutics, Inc.*

Abstract: Binary response experiments are common in scientific studies. However, the study of optimal designs in this area is in a very underdeveloped stage. Sitter and Torsney (1995a) studied optimal designs for binary response experiments with two design variables. In this paper, we consider a general situation with multiple design variables. A novel approach is proposed to identify optimal designs for the commonly used multi-factor logistic and probit models. We give explicit formulas for a large class of optimal designs, including D -, A -, and E -optimal designs. In addition, we identify the general structure of optimal designs, which has a relatively simple format. This property makes it feasible to solve seemingly intractable problems. This result can also be applied in a multi-stage approach.

Key words and phrases: A -optimality, D -optimality, E -optimality, Loewner ordering, logistic model, probit model.

1. Introduction

We consider experiments with a binary response in which a subject is administered m covariates at level $X_i^T = (1, x_{i1}, \dots, x_{im})$. Here X_i^T represents a vector of $m + 1$ design variables selected from a design space $\mathcal{X} \subset R^m$. A typical analysis for this situation is a multi-factor logistic or probit regression model that can be written as

$$Prob(Y_i = 1) = P(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}). \quad (1.1)$$

Here, Y_i is the response of subject i with covariates level X_i , $\beta = (\beta_0, \dots, \beta_m)$ are unknown parameters with $\beta_j \neq 0$ for $j > 0$, and $P(x)$ is a cumulative distribution function (cdf). The two most commonly used $P(x)$'s are $e^x/(1 + e^x)$ for the logistic model and $\Phi(x)$, the cdf of the standard normal distribution, for the probit model. Such models have been studied extensively for data analysis (see for example, Agresti (2002)), but little is known about design selection. With a careful choice of design, statistical inferences can be greatly improved. From a cost-benefit perspective, an efficient design can reduce the sample size needed for achieving a specified precision, or improve the precision for a given sample size.

A complication in studying optimal designs when using a nonlinear model is that, unlike the case of a linear model, the information matrices and optimal designs depend on the unknown parameters. Thus the challenge in designing an experiment for such a model is that one is looking for the best design with the aim of estimating the unknown parameters, and yet one has to know the parameters to find the best design. One way to solve this problem is to use a locally optimal design based on the best guess of the parameters. Other ways are available to address this issue, for example, by using a Bayesian approach (Agin and Chaloner (1999)). While a good guess may not always be available, the locally optimal design approach remains of value. As pointed out in Ford, Torsney, and Wu (1992), locally optimal designs are important if good initial parameter estimates are available from previous experiments. They can also be a benchmark for designs chosen to satisfy experimental constraints. Most of the currently available results pertain to locally optimal designs. Hereafter, the word “locally” is omitted for simplicity.

Many optimality results for GLMs focus on models with one covariate. Ford, Torsney, and Wu (1992) studied c -optimal and D -optimal designs; Sitter and Wu (1993a,b) studied A - and F -optimal designs; Dette and Haines (1994); investigated E -optimal designs. Mathew and Sinha (2001) obtained a series of optimality results for the logistic model by using an algebraic approach, whereas Biedermann, Dette, and Zhu (2006) recently obtained Φ_p -optimal designs for a restricted design space using a geometric approach.

These contributions are important. However, frequently in practice the response is affected by more than one covariate and thus multiple-covariate GLMs are commonly used (Agresti (2002)). Most of the efforts in optimizing designs in this setting rely on limited computational tools. Methodological research on optimal designs is still lacking, and theoretical guidance remains at a very underdeveloped stage. Computational results are mainly achieved by search methods. Notable works include Woods et al. (2006) and Dror and Steinberg (2006) for studying robust designs, as well as Dror and Steinberg (2008) for studying sequential designs. These all used D -optimality and provided algorithms. For example, Dror and Steinberg (2006) provided computer programs for deriving D -optimal designs for general models. We are aware of only three papers that provide explicit formulas in the setting of generalized linear models. Russell et al. (2009) obtained an explicit formula for D -optimal designs under a Poisson regression model, which has the same format of linear predictors as those in Model (1.1). Sitter and Torsney (1995a) applied the geometric approaches of Silvey and Titterton (1973) and of Elfving (1952), respectively, to study D - and c -optimal designs when there are two covariates in Model (1.1). Sitter and Torsney (1995b) extended the results to D -optimal designs when $m > 2$ in Model

(1.1). Under a slightly different set-up, Haines et al. (2007) studied D - optimal designs for logistic regression in two variables. Although the geometric approach is a powerful tool for studying nonlinear designs, it has its limitations: it works fine when the dimension of the parameters is two; it becomes more complicated when the dimension is three, and seems intractable when the dimension is larger than three (Elfving (1952)). Khuri et al. (2006) surveyed design issues for GLMs and pointed out that results on designs for generalized linear models with multiple covariates requires extensive work to evaluate “optimal” or, at least, efficient designs.

Yang and Stufken (2009) proposed an algebraic approach to nonlinear models with two parameters. In this paper, we extend this to optimal designs for GLMs with multiple covariates. With a focus on logistic and probit models, we identify a dominating class of relatively simple designs, which means that for any design ξ that does not belong to this class, there is a design in the class that has an information matrix that dominates ξ in the Loewner ordering. Therefore, we can focus on the subclass when we derive optimal designs. This structural property makes identifying optimal designs for multi-factor GLMs a feasible task. Specifically, we give explicit formulas for all or some of the parameters of a large class of optimal designs. This structural property can also be applied in a multi-stage approach. This is important, because, in a multi-stage approach, the first stage may give us information about the unknown parameters, which can in turn be used in the local optimality approach for adding additional design points in the second stage.

This paper is organized as follows. In Section 2, we introduce the models and the corresponding information matrices. We also identify the structure of optimal designs for GLMs with multiple covariates. Explicit formulas are given in Section 3 for a large class of D -, A -, and E -optimal designs. The parameters of interest can be the full or a subset of the parameters. A closing discussion is presented in Section 4, and the proofs of the technical results given in the Appendix.

2. Statistical Models and Information Matrices

Under (1.1), an exact design can be presented as $\{(X_i, n_i), i = 1, \dots, k\}$, where n_i is the number of subjects with covariates X_i . With n denoting the total number of subjects, $\sum_i n_i = n$. Since finding an optimal exact design is a difficult and often intractable optimization problem, the corresponding approximate design, in which n_i/n is replaced by ω_i , is considered. Thus a design can be given as $\xi = \{(X_i, \omega_i), i = 1, \dots, k\}$, where $\omega_i > 0$ and $\sum_i \omega_i = 1$. For known parameters, there is a one-to-one mapping between X_i and C_i , where

$C_i^T = (1, x_{i1}, \dots, x_{i,m-1}, c_i)$. Here $c_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}$. It is convenient to denote the design ξ as $\xi = \{(C_i, \omega_i), i = 1, \dots, k\}$.

By standard methods, the information matrix for β under (1.1) can be written as

$$\begin{aligned}
 I_\xi(\beta) &= n \sum_{i=1}^k \omega_i X_i \Psi(c_i) (X_i)^T \\
 &= n A(\beta) \left(\sum_{i=1}^k \omega_i C_i \Psi(c_i) (C_i)^T \right) A^T(\beta), \tag{2.1}
 \end{aligned}$$

where $\Psi(x) = [P'(x)]^2/[P(x)(1 - P(x))]$. In deriving (2.1), we utilize $X_i = A(\beta)C_i$. Here $A(\beta) = \begin{pmatrix} I_m & 0 \\ A_1(\beta) & 1/\beta_m \end{pmatrix}$, where $A_1(\beta) = (-\beta_0/\beta_m, -\beta_1/\beta_m, \dots, -\beta_{m-1}/\beta_m)$.

For multi-factor GLMs, $m - 1$ covariates must be bounded, otherwise, the optimality criterion can be made arbitrarily large by the choice of design (Sitter and Torsney (1995a)). Although Dorta-Guerra, Gonzalez-Davina, and Ginebra (2008) showed that bounds are not needed, their conclusion is based on the assumption that the covariates only take two values. Many researchers choose the constraints as $[-1, 1]^m$ (Dror and Steinberg (2006, 2008); Woods et al. (2006)). In this paper, we assume that the first $m - 1$ covariates are bounded, i.e., $x_{ij} \in [U_j, V_j]$, $j = 1, \dots, m - 1$. There is no constraint on the last covariate, i.e., $x_{im} \in (-\infty, \infty)$. In this paper, we show that for an optimal design, all covariates take two values except for one covariate (the one without constraints), which can take 2^m possible values.

Suppose that we are interested in $\eta = F(\beta)$, a vector-valued function of β . For any two designs ξ_1 and ξ_2 , if $I_{\xi_1}(\beta) \leq I_{\xi_2}(\beta)$ (here and elsewhere, matrix inequalities are under the Loewner ordering), then design ξ_2 is at least as good as design ξ_1 for $F(\beta)$ under the commonly used optimality criteria. This can be easily verified by the following equation

$$\Sigma_\xi(\hat{\eta}) = \frac{\partial F(\beta)}{\partial \beta^T} I_\xi^{-1}(\beta) \left(\frac{\partial F(\beta)}{\partial \beta^T} \right)^T. \tag{2.2}$$

Here, $\Sigma_\xi(\hat{\eta})$ is the variance-covariance matrix of $\hat{\eta} = F(\hat{\beta})$, where $\hat{\beta}$ is the MLE of β .

Next, we will show that for any given design $\xi = \{(C_i, \omega_i), i = 1, \dots, k\}$, there exists a design $\tilde{\xi}$ with a simple form such that $I_\xi(\theta) \leq I_{\tilde{\xi}}(\theta)$. Let

$$a_{l,j} = \begin{cases} U_j & \lceil \frac{l}{2^{m-1-j}} \rceil \text{ is odd,} \\ V_j & \lceil \frac{l}{2^{m-1-j}} \rceil \text{ is even,} \end{cases} \quad l = 1, \dots, 2^{m-1}; j = 1, \dots, m - 1, \tag{2.3}$$

where $\lceil a \rceil$ is the smallest integer greater than or equal to a .

Theorem 1. For any given design $\xi = \{(C_i, \omega_i), i = 1, \dots, k\}$, there exists a design $\tilde{\xi}$ such that $I_\xi(\beta) \leq I_{\tilde{\xi}}(\beta)$. Here

$$\tilde{\xi} = \{(\tilde{C}_{l1}, \omega_{l1}) \text{ and } (\tilde{C}_{l2}, \omega_{l2}), l = 1, \dots, 2^{m-1}\}, \quad (2.4)$$

where $(\tilde{C}_{l1})^T = (1, a_{l,1}, \dots, a_{l,m-1}, \tilde{c}_l)$, $(\tilde{C}_{l2})^T = (1, a_{l,1}, \dots, a_{l,m-1}, -\tilde{c}_l)$, and $\tilde{c}_l > 0$; $a_{l,j}$ is either U_j or V_j , and $(a_{l,1}, \dots, a_{l,m-1}), l = 1, \dots, 2^{m-1}$, are all possible such combinations.

Proof. By (2.1) and Lemma 1 in the Appendix, we have

$$I_\xi(\beta) \leq nA(\beta) \left(\sum_{i=1}^k \sum_{l=1}^{2^{m-1}} \omega_i^l C_i^l \Psi(c_i) (C_i^l)^T \right) A^T(\beta), \quad (2.5)$$

where $(C_i^l)^T = (1, a_{l,1}, \dots, a_{l,m-1}, c_i)$, and ω_i^l is the associated weight. Notice that

$$C_i^l \Psi(c_i) (C_i^l)^T = B_l \begin{pmatrix} \Psi(c_i) & c_i \Psi(c_i) \\ c_i \Psi(c_i) & c_i^2 \Psi(c_i) \end{pmatrix} B_l^T, \quad (2.6)$$

where $B_l^T = \begin{pmatrix} 1 & a_{l,1} & \dots & a_{l,m-1} & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$. By (2.5) and (2.6), we have

$$\begin{aligned} I_\xi(\beta) &\leq nA(\beta) \left(\sum_{l=1}^{2^{m-1}} B_l \left(\sum_{i=1}^k \omega_i^l \begin{pmatrix} \Psi(c_i) & c_i \Psi(c_i) \\ c_i \Psi(c_i) & c_i^2 \Psi(c_i) \end{pmatrix} \right) B_l^T \right) A^T(\beta) \\ &\leq nA(\beta) \left(\sum_{l=1}^{2^{m-1}} \sum_{i=1}^2 \omega_{li} \tilde{C}_{li} \Psi(\tilde{c}_l) \tilde{C}_{li}^T \right) A^T(\beta) \\ &= I_{\tilde{\xi}}(\beta). \end{aligned} \quad (2.7)$$

The second inequality in (2.7) is due to Lemma 2 in the Appendix and the fact that $\Psi(\tilde{c}_l) = \Psi(-\tilde{c}_l)$.

Torsney and Gunduz (2001) derived a similar structure for D -optimal designs. With Theorem 1, we can restrict our focus to a subclass to search for optimal designs based on the information matrix. These designs have a relatively simple format: except for the last covariate, all have been identified. Totally there are 2^{m-1} points to be identified when we search for a specific design. When m is small, say $m \leq 3$, we can use a computer algorithm to find them.

However, when m is moderate to large, a computer search is out of the question and the best solution for this situation is an explicit formula, if available.

3. Explicit Formulas for Optimal Designs

In this section, we provide closed-form solutions for a large class of optimal designs. Instead of studying the original parameter, β , directly, we consider the transformed parameter $\theta = (\theta_0, \theta_1, \dots, \theta_m)$. Here $\theta_0 = (\beta_0 + \sum_{j=1}^{m-1} \beta_j(U_j + V_j)/2)/\beta_m$, $\theta_j = \beta_j/\beta_m$, $j = 1, \dots, m - 1$, and $\theta_m = \beta_m$. When the constraints are symmetric ($U_j = -V_j$), for example $[-1, 1]^{m-1}$, then $\theta_0 = \beta_0/\beta_m$. Under the commonly used D -optimality, an optimal design is invariant to such a transformation, and many optimality results have been obtained in this way. Examples are found in Minkin (1987), Sitter and Wu (1993a,b), Sitter and Torsney (1995a), Mathew and Sinha (2001), and Biedermann, Dette, and Zhu (2006), to name a few.

We consider D -, A -, and E -optimality, the three commonly used optimality criteria. A D -optimal design maximizes the determinant of the information matrix and minimizes the joint confidence ellipsoid of the parameters. An A -optimal design minimizes the trace of the inverse of the information matrix and minimizes the sum of the variances of parameter estimators. An E -optimal design maximizes the smallest eigenvalue of the information matrix and protects against the worst scenario when we estimate the parameters.

When our main concern is the covariate effects only, optimality results for θ are no longer optimal. We need to study the corresponding information matrix for $\theta^1 = (\theta_1, \dots, \theta_m)$. Since θ^1 is a function of β_1, \dots, β_m only, we consider optimal designs for θ and θ^1 separately.

3.1 Optimal designs for θ

For a given design $\xi = \{(C_i, \omega_i), i = 1, \dots, k\}$, the information matrix for θ can be written as

$$I_\xi(\theta) = n \sum_{i=1}^k \omega_i \tilde{C}_i \Psi(c_i) (\tilde{C}_i)^T, \tag{3.1}$$

where $(\tilde{C}_i)^T = (\theta_m, \theta_m(x_1 - (U_1 + V_1)/2), \dots, \theta_m(x_{m-1} - (U_{m-1} + V_{m-1})/2), c_i/\theta_m)$.

Theorem 2. *Under Model (1.1) for the logistic or probit model, ξ^* is a D -, A -, or E -optimal design for θ if $\xi^* = \{(C_{l1}^*, 1/2^m) \& (C_{l2}^*, 1/2^m), l = 1, \dots, 2^{m-1}\}$, where $(C_{l1}^*)^T = (1, a_{l,1}, \dots, a_{l,m-1}, c^*)$ and $(C_{l2}^*)^T = (1, a_{l,1}, \dots, a_{l,m-1}, -c^*)$, $a_{l,j}$*

is defined at (2.3), and c^* minimizes $f(c)$, with

$$f(c) = \begin{cases} c^{-2}(\Psi(c))^{-m-1}, & D\text{-optimality;} \\ \beta_m^2(c^2\Psi(c))^{-1} + \frac{1}{\beta_m^2} \left(1 + \sum_{j=1}^{m-1} \frac{4}{(V_j-U_j)^2} \right) (\Psi(c))^{-1}, & A\text{-optimality;} \\ \max\{(\beta_m^2\Psi(c))^{-1}, (\frac{1}{4}\beta_m^2(V_1-U_1)^2\Psi(c))^{-1}, \dots, \\ (\frac{1}{4}\beta_m^2(V_{m-1}-U_{m-1})^2\Psi(c))^{-1}, (\frac{1}{\beta_m^2}c^2\Psi(c))^{-1}\}, & E\text{-optimality.} \end{cases} \tag{3.2}$$

Proof. The proofs for D -, A -, and E -optimal designs are completely analogous, so we prove only the result for D -optimality. First, we can limit our considerations to a design such that its information matrix is positive definite, otherwise θ is not estimable. By Theorem 1 and (2.2), for any such design ξ , there exists a design $\tilde{\xi}$ defined in (2.4) with $I_\xi(\theta) \leq I_{\tilde{\xi}}(\theta)$. Thus we have $Det(I_\xi(\theta)) \leq Det(I_{\tilde{\xi}}(\theta))$. By (3.1) and (2.4), the (i, j) 'th ($i \leq j$) element of $I_{\tilde{\xi}}(\theta)$ is

$$\left\{ \begin{array}{ll} \theta_m^2 \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \Psi(\tilde{c}_l), & i = j = 1; \\ \frac{1}{4} \theta_m^2 (V_{j-1} - U_{j-1})^2 \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \Psi(\tilde{c}_l), & i = j, j = 2, \dots, m; \\ \frac{1}{\theta_m^2} \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \tilde{c}_l^2 \Psi(\tilde{c}_l), & i = j = m + 1; \\ \theta_m^2 \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) (a_{l,j-1} - \frac{U_{j-1} + V_{j-1}}{2}) \Psi(\tilde{c}_l), & i = 1, j = 2, \dots, m; \\ \sum_{l=1}^{2^{m-1}} (\omega_{l1} - \omega_{l2}) \tilde{c}_l \Psi(\tilde{c}_l), & i = 1, j = m + 1; \\ \theta_m^2 \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) (a_{l,i-1} - \frac{U_{i-1} + V_{i-1}}{2}) (a_{l,j-1} - \frac{U_{j-1} + V_{j-1}}{2}) \Psi(\tilde{c}_l), & i \neq j, i, j = 2, \dots, m; \\ \sum_{l=1}^{2^{m-1}} (\omega_{l1} - \omega_{l2}) (a_{l,j-1} - \frac{U_{j-1} + V_{j-1}}{2}) \tilde{c}_l \Psi(\tilde{c}_l), & i = m + 1, j = 2, \dots, m. \end{array} \right. \tag{3.3}$$

By (A.4) of Lemma 3 in the Appendix,

$$\begin{aligned}
 \text{Det}(I_{\tilde{\xi}}(\theta)) &\leq \left(\theta_m^2 \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \Psi(\tilde{c}_l) \right) \left(\frac{1}{\theta_m^2} \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \tilde{c}_l^2 \Psi(\tilde{c}_l) \right) \\
 &\quad \prod_{j=1}^{m-1} \left(\frac{1}{4} \theta_m^2 (V_j - U_j)^2 \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \Psi(\tilde{c}_l) \right) \\
 &= \left(\sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \Psi(\tilde{c}_l) \right)^m \left(\sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \tilde{c}_l^2 \Psi(\tilde{c}_l) \right) \\
 &\quad \prod_{j=1}^{m-1} \left(\frac{1}{4} \theta_m^2 (V_j - U_j)^2 \right), \tag{3.4}
 \end{aligned}$$

with equality when all the off-diagonal components of $I_{\tilde{\xi}}(\theta)$ are zeros. By (A.8) of Lemma 4, there exists a point \tilde{c} , such that

$$\begin{aligned}
 \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \Psi(\tilde{c}_l) &= \Psi(\tilde{c}), \\
 \sum_{l=1}^{2^{m-1}} (\omega_{l1} + \omega_{l2}) \tilde{c}_l^2 \Psi(\tilde{c}_l) &\leq \tilde{c}^2 \Psi(\tilde{c}). \tag{3.5}
 \end{aligned}$$

By (3.4) and (3.5), we further have

$$\begin{aligned}
 \text{Det}(I_{\tilde{\xi}}(\theta)) &\leq \Psi(\tilde{c})^m (\tilde{c}^2 \Psi(\tilde{c})) \prod_{j=1}^{m-1} \left(\frac{1}{4} \theta_m^2 (V_j - U_j)^2 \right) \\
 &\leq \Psi(c^*)^m ((c^*)^2 \Psi(c^*)) \prod_{j=1}^{m-1} \left(\frac{1}{4} \theta_m^2 (V_j - U_j)^2 \right), \tag{3.6}
 \end{aligned}$$

the last inequality being due to the fact that c^* minimizes $f(c)$. This is equivalent to maximizing $c^2(\Psi(c))^{m+1}$. On the other hand, applying (3.3), we can directly check that $I_{\xi^*}(\theta)$ is a diagonal matrix, which has the (i, i) 'th diagonal element $\theta_m^2 \Psi(c^*)$ when $i = 1$, $(1/4)\theta_m^2 (V_{j-1} - U_{j-1})^2 \Psi(c^*)$ when $i = 2, \dots, m$, and $(1/\theta_m^2)(c^*)^2 \Psi(c^*)$ when $i = m + 1$. Thus we have

$$\text{Det}(I_{\xi^*}(\theta)) = \Psi(c^*)^m ((c^*)^2 \Psi(c^*)) \prod_{j=1}^{m-1} \left(\frac{1}{4} \theta_m^2 (V_j - U_j)^2 \right). \tag{3.7}$$

Our conclusion follows from (3.7) and (3.6).

Table 1. c^* for logistic and probit models with D -optimality.

m	2	3	4	5	6	7	8
<i>Logistic</i>	1.2229	1.0436	0.9254	0.8399	0.7744	0.7222	0.6793
<i>Probit</i>	0.9376	0.8159	0.7320	0.6696	0.6209	0.5815	0.5487

From the proof of Theorem 2, it is clear that similar conclusions hold for Φ_p -optimality. Theorem 2 provides explicit forms for optimal designs under commonly used optimality criteria. The only thing we need to do is to determine the value of c^* under different optimality criteria. Optimal designs have 2^m support points with the same weights; the 2^m support points have m dimensions, the i th element ($i \leq m - 1$) is either the lower bound U_i or the upper bound V_i , the m 'th element is either c^* or $-c^*$. Notice that a transformation is necessary if we transform the design point C_i to the original design point X_i (see Section 2).

One advantage of D -optimality is that the D -optimal design is invariant under one-to-one parameter transformations. Since θ is such a transformation of β , the D -optimal design we obtained here is also D -optimal for the original parameter β .

Corollary 1. *Under Model (1.1), for the logistic or probit link function, ξ^* is a D -optimal design for parameter β . Here, $\xi^* = \{(C_{l1}^*, 1/2^m) \& (C_{l2}^*, 1/2^m), l = 1, \dots, 2^{m-1}\}$, where $(C_{l1}^*)^T = (1, a_{l,1}, \dots, a_{l,m-1}, c^*)$ and $(C_{l2}^*)^T = (1, a_{l,1}, \dots, a_{l,m-1}, -c^*)$, $a_{l,j}$ is defined in (2.3), and c^* maximizes $c^2(\Psi(c))^{m+1}$.*

Under D -optimality, the value of c^* only depends on the value of m , the number of covariates. Although it is easy to compute the value of c^* , for convenience we table some of its values.

Sitter and Torsney (1995a,b) also derived D -optimal designs for θ under Model (1.1). Corollary 1 is consistent with the Sitter and Torsney (1995a) D -optimality results for logistic and probit models under (1.1) when $m = 2$, and with the Sitter and Torsney (1995b) D -optimality results (Table 2) for logistic models under (1.1) when $m = 8$. Dror and Steinberg (2006) provide an algorithm and the corresponding program to derive D -optimality designs. The program can identify an exact D -optimal design when m is moderate, say less than 5. However, when m gets larger, the result is not optimal, but remains highly efficient.

Under A - or E - optimality, optimal designs depend on the values of β_m and on the restricted regions $[U_i, V_i]$. Given these values, we can derive the value of c^* by minimizing $f(c)$ in (3.2). Here is an example to illustrate the application, with $m = 3$ and $[U_i, V_i] = [-1, 1]$, $i = 1, 2$.

For an A -optimal design, we need to find the value of c^* that minimizes $\beta_3^2(c^2\Phi(c))^{-1} + 3\Phi^{-1}(c)/\beta_3^2$. Assume $\beta_3 = 1$. By routine algebra, we find $c^* = 1.0238$ for the logistic link function and 0.8874 for the probit link function, but

when $\beta_3 = 6$, $c^* = 2.3778$ for the logistic link function and 1.5709 for the probit link function, respectively.

For E -optimal designs, we need to find the value of c^* that minimizes $\text{Max}\{(\beta_3^2\Phi(c))^{-1}, \beta_3^2(c^2\Phi(c))^{-1}\}$. Interestingly, c^* takes two values only. For the logistic model, $c^* = \beta_3^2$ when $\beta_3 \leq 1.549$, and is 2.3994 otherwise. For the probit model, $c^* = \beta_3^2$ when $\beta_3 \leq 1.255$, and is 1.575 otherwise.

3.2 Optimal designs for θ^1

We first need $I_\xi(\theta^1)$, the information matrix for θ^1 . Rewrite $I_\xi(\theta)$ in (3.1) as $\begin{pmatrix} I_{11} & I_{12} \\ I_{12}' & I_{22} \end{pmatrix}$, where I_{11} is a scalar, I_{12} is $1 \times m$ vector, I_{22} is $m \times m$ matrix, so $I_\xi(\theta^1) = I_{22} - I_{12}'I_{11}^{-1}I_{12}$. Clearly, $I_\xi(\theta^1) \leq I_{22}$ with equality when $I_{12}'I_{11}^{-1}I_{12} = 0$. A special case of the latter occurs if $I_\xi(\theta)$ is a diagonal matrix. Applying the same argument to I_{22} as that used on $I_\xi(\theta)$ in Theorem 2, we have the following.

Theorem 3. *Under Model (1.1), for the logistic or probit model, ξ^* is a D -, A -, or E -optimal design for parameter θ^1 if $\xi^* = \{(C_{l1}^*, 1/2^m) \& (C_{l2}^*, 1/2^m), l = 1, \dots, 2^{m-1}\}$, where $(C_{l1}^*)^T = (1, a_{l,1}, \dots, a_{l,m-1}, c^*)$ and $(C_{l2}^*)^T = (1, a_{l,1}, \dots, a_{l,m-1}, -c^*)$, $a_{l,j}$ is defined in (2.3), and c^* minimizes $f^1(c)$ with*

$$f^1(c) = \begin{cases} c^{-2}(\Psi(c))^{-m}, & D\text{-optimality;} \\ \beta_m^2(c^2\Psi(c))^{-1} + \frac{1}{\beta_m^2} \left(\sum_{j=1}^{m-1} \frac{4}{(V_j - U_j)^2} \right) (\Psi(c))^{-1}, & A\text{-optimality;} \\ \max\left\{ \left(\frac{1}{4}\beta_m^2(V_1 - U_1)^2\Psi(c) \right)^{-1}, \dots, \right. \\ \quad \left. \left(\frac{1}{4}\beta_m^2(V_{m-1} - U_{m-1})^2\Psi(c) \right)^{-1}, \left(\frac{1}{\beta_m^2}c^2\Psi(c) \right)^{-1} \right\}, & E\text{-optimality.} \end{cases} \tag{3.8}$$

Using the same reasoning, the conclusion in Theorem 3 can be extended to other subsets of θ with suitable $f^1(c)$. Since θ^1 is a one-to-one transformation of $(\beta_1, \dots, \beta_m)$, the invariance of the D -optimality criterion yields the following.

Corollary 2. *Under Model (1.1), for the logistic and probit link functions, ξ^* is a D -optimal design for parameter $(\beta_1, \dots, \beta_m)$. Here, $\xi^* = \{(C_{l1}^*, 1/2^m) \text{ and } (C_{l2}^*, 1/2^m), l = 1, \dots, 2^{m-1}\}$, where $(C_{l1}^*)^T = (1, a_{l,1}, \dots, a_{l,m-1}, c^*)$ and $(C_{l2}^*)^T = (1, a_{l,1}, \dots, a_{l,m-1}, -c^*)$, $a_{l,j}$ is defined in (2.3), and c^* maximizes $c^2(\Psi(c))^m$.*

Titterington (1978) proved Corollary 2 for linear regression models. This result was extended by Martin-Martin, Torsney, and Lopez-Fidalgo (2007) with respect to marginally and conditionally restricted designs.

Notice that c^* in Corollary 1 maximizes $c^2(\Psi(c))^{m+1}$, thus the value of c^* in Corollary 2 is the same as the value of c^* in Corollary 1 for $m - 1$. For example, when $m = 3$, c^* is the same as the c^* in Corollary 1 when $m = 2$, as given in Table 1. When $m = 2$, c^* is 1.5434 and 1.1381 for the logistic and probit link functions, respectively.

Corollary 2 gives D -optimal designs for $(\beta_1, \dots, \beta_m)$ excluding the intercept parameter β_0 . This is particularly useful in some practical situations where the main interest is in the coefficients of covariates; it applies as well to the situation where the main interest is in estimating some of the coefficients, excluding the constant term.

For A - and E -optimal designs, the value of c^* depends on the parameters and the restricted regions. Given these values, it is straightforward to compute c^* . Notice that the information matrices of the optimal designs in Theorem 3 are diagonal. Thus all estimators are uncorrelated and the optimal designs are orthogonal.

3.3. Optimal designs based on a subset of 2^m points

Although Theorems 2 and 3 give explicit formulas for optimal designs for θ and θ^1 , the number of support points grows quickly as m increases. For example when m is 8, one needs 256 support points. Sitter and Torsney (1995b), however, note that a D -optimal design can be based on a subset of 2^m design points. Their idea utilizes a nice property of Hadamard matrices and they provide a D -optimal design based on 16 points for $m = 8$.

Applying this idea, the optimal designs given in Theorems 2 and 3 can also be based on less than 2^m points. The procedure can be described as follows: (i) generate a $k \times (m + 1)$ ($k \geq m + 1$) matrix by selecting any $m + 1$ columns of a $k \times k$ Hadamard matrix including the column with all 1's; (ii) (a) leave the column of all 1's unchanged; (b) in each of the $m - 1$ columns that corresponds to one covariate $x_{i,j}$, re-label the upper bound V_j as 1 and the lower bound U_j as -1; (c) in the remaining column that corresponds to the induced covariate c_i , re-label "1" as c^* and "-1" as $-c^*$. Then the derived design can be based on k support points, each row of the resulting matrix being a support point $(1, x_{i1}, \dots, x_{i,m-1}, c_i)$ with weight $1/k$. It can be verified that the derived design has the same information matrix as the design ξ^* given in Theorems 2 and 3 and thus is optimal. A $k \times k$ Hadamard matrix exists as long as $k = 4k_1$ for any positive integer k_1 less than 100. We illustrate the procedure for $m = 7$ by giving a D -optimal design for θ^1 under Model (1.1) with Logistic link function. Notice that each point has weight $1/8$.

design point i	1	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$x_{i,4}$	$x_{i,5}$	$x_{i,6}$	c_i
1	1	V_1	V_2	V_3	V_4	V_5	V_6	0.7744
2	1	V_1	V_2	U_3	U_4	U_5	V_6	-0.7744
3	1	V_1	U_2	V_3	U_4	U_5	U_6	0.7744
4	1	V_1	U_2	U_3	V_4	V_5	U_6	-0.7744
5	1	U_1	V_2	V_3	V_4	U_5	U_6	-0.7744
6	1	U_1	V_2	U_3	U_4	V_5	U_6	0.7744
7	1	U_1	U_2	V_3	U_4	V_5	V_6	-0.7744
8	1	U_1	U_2	U_3	V_4	U_5	V_6	0.7744

4. Discussion

Although multi-factor GLMs are widely applied in practice, the research on optimal designs is very limited. This paper provides a step in the search for optimal designs for GLMs with multiple covariates.

To our knowledge, there are few optimality results available for a subset of parameters; we are only aware of the c -optimality result in Sitter and Torsney (1995a). There are no optimality results available for multi-factor GLMs under A - or E -optimality. This paper provides explicit solutions for these questions.

An interesting question concerns optimal designs when there are interactions among the covariates under Model (1.1). Both Woods et al. (2006) and Dror and Steinberg (2006) considered this situation but is not clear whether there exists a general structure for this type of optimal design. Because of the interaction terms, it seems that a strategy similar to that of Theorem 1 cannot be applied.

Acknowledgement

We received numerous constructive and helpful comments from the Editor, an associate editor, and two referees. These comments and suggestions substantially changed and improved the article. We thank them. The research of Min Yang was supported by NSF grants DMS-0707013 and DMS-0748409.

Appendix

Lemma 1. For any design point (C_i, ω_i) , there exist weights ω_i^l , $l = 1, \dots, 2^{m-1}$, such that

$$\omega_i C_i \Psi(c_i) (C_i)^T \leq \sum_{l=1}^{2^{m-1}} \omega_i^l C_i^l \Psi(c_i) (C_i^l)^T. \quad (\text{A.1})$$

Here $(C_i^l)^T = (1, a_{l,1}, \dots, a_{l,m-1}, c_i)$ and $\sum_{l=1}^{2^{m-1}} \omega_i^l = \omega_i$.

Proof. Let $r_j = (V_j - x_{i,j})/(V_j - U_j)$, $j = 1, \dots, m - 1$. Then it is easy to show that

$$\begin{aligned} r_j U_j + (1 - r_j) V_j &= x_{i,j}, \\ r_j U_j^2 + (1 - r_j) V_j^2 &\geq x_{i,j}^2. \end{aligned} \tag{A.2}$$

Now we consider two points $C_{i,1} = (1, U_1, x_{i,2}, \dots, x_{i,m-1}, c_i)'$ and $C_{i,2} = (1, V_1, x_{i,2}, \dots, x_{i,m-1}, c_i)'$ with corresponding weights $\omega_{i,1} = r_1 \omega_i$ and $\omega_{i,2} = (1 - r_1) \omega_i$, respectively. The matrices $\omega_{i,1} C_{i,1} \Psi(c_i) (C_{i,1})^T + \omega_{i,2} C_{i,2} \Psi(c_i) (C_{i,2})^T$ and $\omega_i C_i \Psi(c_i) (C_i)^T$ have the same elements except for their second diagonal elements; the former is greater than the latter according to (A.2). Thus we have

$$\omega_{i,1} C_{i,1} \Psi(c_i) (C_{i,1})^T + \omega_{i,2} C_{i,2} \Psi(c_i) (C_{i,2})^T \geq \omega_i C_i \Psi(c_i) (C_i)^T.$$

For the point $C_{i,1}$, using the same argument for $x_{i,2}$ as used for $x_{i,1}$ with C_i , we can improve the information matrix by replacing $C_{i,1}$ with two new points $C_{i,1,1}$ and $C_{i,1,2}$, that are the same as $C_{i,1}$ except that $x_{i,2}$ is replaced by U_2 and V_2 , respectively. Similarly we generate two new points $C_{i,2,1}$ and $C_{i,2,2}$ from $C_{i,2}$. Then (A.1) can be established by repeating this procedure until $x_{i,m-1}$ is replaced by U_{m-1} or V_{m-1} .

Lemma 2. For any k points $\{(c_i, \omega_i), i = 1, \dots, k\}$, $k \geq 2$, there exist a point c^* and $0 \leq \omega^* \leq \sum_{i=1}^k \omega_i$ such that

$$\begin{aligned} &\sum_{i=1}^k \omega_i \begin{pmatrix} \Psi(c_i) & c_i \Psi(c_i) \\ c_i \Psi(c_i) & c_i^2 \Psi(c_i) \end{pmatrix} \\ &\leq \omega^* \begin{pmatrix} \Psi(c^*) & c^* \Psi(c^*) \\ \Psi(c^*) & (c^*)^2 \Psi(c^*) \end{pmatrix} + \left(\sum_{i=1}^k \omega_i - \omega^* \right) \begin{pmatrix} \Psi(-c^*) & -c^* \Psi(-c^*) \\ -c^* \Psi(-c^*) & (-c^*)^2 \Psi(-c^*) \end{pmatrix}. \end{aligned} \tag{A.3}$$

Here, $\Psi(x) = [P'(x)]^2/[P(x)(1 - P(x))]$, and $P(x)$ is the cumulative distribution function for the logistic or the probit model.

Proof. This is an immediate consequence of Theorem 1 of Yang and Stufken (2009).

Lemma 3. Let A be an $m \times m$ positive matrix and μ_i , $i = 1, \dots, m$, be the eigenvalues of A . Then

$$\prod_{i=1}^m \mu_i \leq \prod_{i=1}^m A_{ii}, \tag{A.4}$$

$$\sum_{i=1}^m \mu_i^{-p} \geq \sum_{i=1}^m A_{ii}^{-p}, p > 0, \tag{A.5}$$

$$\min\{\mu_1, \dots, \mu_m\} \leq \min\{A_{11}, \dots, A_{mm}\}, \tag{A.6}$$

where, A_{ii} , $i = 1, \dots, m$, are the diagonal elements of the matrix A . If A is a diagonal matrix, then (A.4), (A.5), and (A.6) are equalities.

Proof. There exists an orthogonal matrix P such that $A = P \text{diag}(\mu_1, \dots, \mu_m) P^T$ and $PP^T = I$. Thus, $A_{ii} = \sum_{j=1}^m P_{ij}^2 \mu_j$, $i = 1, \dots, m$, and $\sum_{j=1}^m P_{ij}^2 = 1$. Immediately we have (A.6). On the other hand, for any convex function $f(x)$, we have

$$\begin{aligned} \sum_{i=1}^m f(A_{ii}) &= \sum_{i=1}^m f\left(\sum_{j=1}^m P_{ij}^2 \mu_j\right) \leq \sum_{i=1}^m \sum_{j=1}^m P_{ij}^2 f(\mu_j) \\ &= \sum_{j=1}^m f(\mu_j) \left(\sum_{i=1}^m P_{ij}^2\right) = \sum_{j=1}^m f(\mu_j). \end{aligned} \quad (\text{A.7})$$

(A.4) and (A.5) follows from (A.7) by taking $f(x) = -\log(x)$ and x^{-p} , $p > 0$, respectively. It is easy to see that the three equality signs hold when A is a diagonal matrix.

Lemma 4. For any k points $\{(c_i, \omega_i), i = 1, \dots, k\}$, where $c_i \geq 0$, $\omega_i \geq 0$, and $\sum_{i=1}^k \omega_i = 1$, there exists \tilde{c} such that

$$\begin{aligned} \sum_{i=1}^k \omega_i \Psi(c_i) &= \Psi(\tilde{c}), \\ \sum_{i=1}^k \omega_i c_i^2 \Psi(c_i) &\leq \tilde{c}^2 \Psi(\tilde{c}), \end{aligned} \quad (\text{A.8})$$

where $\Psi(x) = [P'(x)]^2 / [P(x)(1 - P(x))]$, and $P(x)$ is the cumulative distribution function for the logistic or the probit model.

Proof. Let $\Psi_1(x) = \Psi(x)$ and $\Psi_3(x) = x^2 \Psi(x)$, then check that $\Psi_1(x)$ and $\Psi_3(x)$ satisfy the condition of Proposition A.2. of Yang and Stufken (2009). The conclusion follows an application of that proposition.

References

- Agin, M. and Chaloner, K. (1999). Optimal Bayesian design for a logistic regression model: geometric and algebraic approaches. *Multivariate Analysis, Design of Experiments and Survey Sampling* (Edited by S. Ghosh), 609-624. Dekker, New York.
- Agresti, A. (2002). *Categorical Data Analysis*. 2nd edition. Wiley, New York.
- Biedermann, S., Dette, H. and Zhu, W. (2006). Geometric construction of optimal designs for dose-reponse models with two parameters. *J. Amer. Statist. Assoc.* **101**, 747-759.

- Dette, H. and Haines, L. M. (1994). E -optimal designs for linear and nonlinear models with two parameters. *Biometrika* **81**, 739-754.
- Dorta-Guerra, R., Gonzalez-Davina, E. and Ginebra, J. (2008). Two-level experiments for binary response data. *Comput. Statist. Data Anal.* **53**, 196-208.
- Dror, H. and Steinberg, D. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520-529.
- Dror, H. and Steinberg, D. (2008). Sequential experimental designs for generalized linear models. *J. Amer. Statist. Assoc.* **103**, 288-298.
- Elfving, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23**, 255-262.
- 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. *J. Roy. Statist. Soc. Ser. B* **54**, 569-583.
- Haines, L. M., Kabera, M. G., Ndlovu, P. and O'Brien, T. E. (2007). D -optimal designs for logistic regression in two variables. mODa 8 - Advances in Model-Oriented Design and Analysis (Edited by J. L. LopezFidalgo, J. M. RodriguezDiaz and B. Torsney), 91-98.
- 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.
- Martin-Martin, R., Torsney, B. and Lopez-Fidalgo, J. (2007). Construction of marginally and conditionally restricted designs using multiplicative algorithms. *Comput. Statist. Data Anal.* **12**, 5547-5561.
- Mathew, T. and Sinha, B. K. (2001). Optimal designs for binary data under logistic regression. *J. Statist. Plann. Inference* **93**, 295-307.
- Minkin, S. (1987). Optimal designs for binary data. *J. Amer. Statist. Assoc.* **82**, 1098-1103
- Russell, K., Woods, D., Lewis, S. and Eccleston, J. (2009). D -optimal designs for Poisson regression models. *Statist. Sinica* **19**, 721-730.
- Silvey, S. D. and Titterton, D. M. (1973). A geometric approach to optimal design theory. *Biometrika* **60**, 21-32.
- Sitter, R. R. and Torsney, B. (1995a). Optimal designs for binary response experiments with two design variables. *Statist. Sinica* **5**, 405-419.
- Sitter, R. R. and Torsney, B. (1995b). Optimal designs for generalised linear models. *Proceedings of 4th International Workshop on Model Oriented Data Analysis*, 87-102.
- Sitter, R. R. and Wu, C. F. J. (1993a). Optimal designs for binary response experiments: Fieller, D , and A criteria. *Scand. J. Statist.* **20**, 329-341.
- Sitter, R. R. and Wu, C. F. J. (1993b). On the accuracy of Fieller intervals for binary response data. *J. Amer. Statist. Assoc.* **88**, 1021-1025.
- Titterton, D. M. (1978). Estimation of correlation coefficients by ellipsoidal trimming. *J. Roy. Statist. Soc. Ser. C* **27**, 227-234.
- Torsney, B. and Gunduz, N. (2001). On Optimal Designs for High- Dimensional Binary Regression Models. In *Optimum Design 2000* (Edited by A. C. Atkinson, B. Bogacka and A. A. Zhiglavskii) Kluwer Academic, Boston, 275-286.
- Woods, D., Lewis, S., Eccleston, J. and Russell, K. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48**, 284-292.
- Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* **37**, 518-541.

Department of Statistics, University of Missouri, Columbia, MO 65211-6100, USA.

E-mail: yangmi@missouri.edu

Department of Biostatistics, University of Alabama-Birmingham, Birmingham, Alabama 35294, USA.

E-mail: binzhang@uab.edu

Department of Statistics, Precision Therapeutics, Inc., Pittsburgh, PA 15203, USA.

E-mail: shu444@gmail.com

(Received May 2009; accepted March 2010)