Optimal designs for binary data under logistic regression

Thomas Mathew[a,*], Bikas Kumar Sinha[b,1]

[a] *Department of Mathematics and Statistics, University of Maryland, 1000 Hilltop Circle, Baltimore, MD 21250, USA*

[b] *Indian Statistical Institute, 203 B.T. Road, Calcutta 700035, India*

Received 5 November 1999; received in revised form 15 May 2000; accepted 21 July 2000

Abstract

A unified approach is presented for the derivation of D- and A-optimal designs for binary data under the two-parameter logistic regression model. The optimal design is constructed for the estimation of several pairs of parameters. The E-optimal design is also obtained in some cases.

MSC: 62J02; 62K05

Keywords: A-optimality; D-optimality; E-optimality; Information matrix; Weak supermajorization

1. Introduction

Consider a binary response Y_x resulting from a non-stochastic dose level x. Assume that Y_x takes the values 0 and 1 and the probability that Y_x takes the value 1 is given by

$$P(x) = \frac{1}{1 + e^{-(\alpha+\beta x)}}, \tag{1.1}$$

where α and β are unknown parameters with $\beta > 0$. Consider m distinct dose levels $x_1, x_2, \ldots, x_m$, and suppose we wish to obtain n_i observations on Y at dose level x_i ($i = 1, 2, \ldots, m$). Let $\sum_{i=1}^{m} n_i = n$. For the estimation of α and β, or some functions of α and β, the optimal design problem in this context consists of optimally selecting the x_i's (in a given region) and the n_i's, with respect to some optimality criterion, for a fixed n. The estimation problems that are usually of interest refer to (a) the estimation of β, or α/β, or some percentile of $P(x)$ given in (1.1), or (b) the joint estimation of a pair of parameters such as (i) α and β, (ii) β and α/β, (iii) β and a percentile of $P(x)$, and (iv) two percentiles of $P(x)$. For estimating any of the individual parameters given above, we can consider the asymptotic variance of the maximum likelihood estimator, and then choose the x_i's and n_i's optimally by minimizing this asymptotic variance. For the joint estimation of two parameters, we can consider the information matrix of the two parameters and then choose the x_i's and the n_i's to minimize a suitable scalar-valued function of the information matrix. This amounts to minimizing suitable scalar-valued functions of the asymptotic variance–covariance matrix of the maximum likelihood estimators of the parameters. The D- and A-optimality criteria are well-known examples. Some of the relevant references on this specific optimal design problem include Abdelbasit and Plackett (1983), Minkin (1987), Khan and Yazdi (1988), Wu (1988), Ford et al. (1992), Sitter and Wu (1993) and Hedayat et al. (1997). While the D-optimality criterion has received considerable attention in this context, A-optimality has also been considered by some authors (see Sitter and Wu, 1993). The optimum dose levels actually depend on the unknown parameters α and β, as is typical in non-linear settings. In fact, solutions to the optimal design problems mentioned above provide optimum values of $\alpha+\beta x_i$. Hence, in order to implement the design in practice, good initial estimates of α and β must be available. In spite of this unpleasant feature, it is important to construct the optimal designs in this context; see the arguments in Ford et al. (1992, p. 569).

In the present article, we consider the joint estimation of (i) α and β, (ii) β and α/β, (iii) β and a percentile of $P(x)$, and (iv) two percentiles of $P(x)$, and provide a unified approach for the construction of D- and A-optimal designs. It should be noted that if $I(\alpha,\beta)$ denotes the information matrix of (α,β) and if θ_1 and θ_2 are two functions of α and β, then the information matrix of (θ_1,θ_2) is $JI(\alpha,\beta)J'$, where the matrix J does not depend on the dose levels. Hence, the D-optimal design is the same for estimating any two functions of α and β. Even though we have constructed mainly the D- and A-optimal designs, we have been able to derive the E-optimal design in some special cases. Most of the time, we have been able to derive the A-optimal design only within the class of symmetric designs, i.e., the class of designs where the dose levels x_i are such that both $\alpha+\beta x_i$ and $-(\alpha+\beta x_i)$ occur with equal weights.

For simplicity, we shall consider the continuous setting in which n_i/n is replaced by ξ_i, where the ξ_i's satisfy $\xi_i > 0$ and $\sum_{i=1}^{m}\xi_i = 1$. Thus, a design can be denoted by $\mathscr{D} = \{(x_i,\xi_i),\ i = 1,2,\ldots,m\}$. We tacitly assume the dose region to be $0 < x < \infty$. As a general reference on optimal designs, we have used Pukelsheim (1993) as and when necessary. In particular, for Lemma 1 in Section 2, we have borrowed clues from Pukelsheim (1993, Section 10.5).

2. Estimation of α and β

The following lemma will play a crucial role in our derivation of optimal designs.

Lemma 1. *Let ξ_i's $(i = 1,2,\ldots,m)$ be positive real numbers satisfying $\sum_{i=1}^m \xi_i = 1$. For any given set of distinct real numbers $a_1, a_2, \ldots, a_m$, there exists c satisfying*

$$\sum_{i=1}^m \xi_i \frac{e^{a_i}}{(1+e^{a_i})^2} = \frac{e^c}{(1+e^c)^2}, \tag{2.1}$$

$$\sum_{i=1}^m \xi_i a_i^2 \frac{e^{a_i}}{(1+e^{a_i})^2} \leqslant c^2 \frac{e^c}{(1+e^c)^2}. \tag{2.2}$$

The proof of the lemma is given in the appendix.

Let

$$a_i = \alpha + \beta x_i. \tag{2.3}$$

Using the notation ξ_i instead of n_i/n, the information matrix for the joint estimation of α and β is

$$I(\alpha,\beta) = \begin{pmatrix} \sum_{i=1}^m \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2} & \sum_{i=1}^m \xi_i x_i \frac{e^{-a_i}}{(1+e^{-a_i})^2} \\ \sum_{i=1}^m \xi_i x_i \frac{e^{-a_i}}{(1+e^{-a_i})^2} & \sum_{i=1}^m \xi_i x_i^2 \frac{e^{-a_i}}{(1+e^{-a_i})^2} \end{pmatrix}, \tag{2.4}$$

where a_i is given in (2.3). Note that for any real number a, $e^a/(1+e^a)^2 = e^{-a}/(1+e^{-a})^2$. This is a property that we shall frequently use.

2.1. D-*optimality*

The solution to the D-optimal design problem is already available in the literature (see Minkin, 1987; Khan and Yazdi, 1988; Sitter and Wu, 1993). We shall give a simple derivation, applying Lemma 1. We need to maximize $|I(\alpha,\beta)|$ in order to obtain the D-optimal design. Note that

$$\beta^2 |I(\alpha,\beta)| = \left[\sum_{i=1}^m \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right] \left[\sum_{i=1}^m \xi_i a_i^2 \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right] - \left[\sum_{i=1}^m \xi_i a_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right]^2. \tag{2.5}$$

It is readily seen that a D-optimal design should be symmetric in the a_i's, i.e., both a_i and $-a_i$ should occur with the same weight. For such a symmetric design, (2.5) simplifies to

$$\beta^2 |I(\alpha,\beta)| = \sum_{i=1}^m \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2} \sum_{i=1}^m \xi_i a_i^2 \frac{e^{-a_i}}{(1+e^{-a_i})^2}, \tag{2.6}$$

where we have also used the fact that $e^{-a_i}/(1+e^{-a_i})^2 = e^{a_i}/(1+e^{a_i})^2$. In view of Lemma 1 and (2.6), for any symmetric design, there exists c satisfying

$$\beta^2 |I(\alpha,\beta)| \leqslant \frac{e^c}{(1+e^c)^2} c^2 \frac{e^c}{(1+e^c)^2} = c^2 \frac{e^{2c}}{(1+e^c)^4}. \tag{2.7}$$

In other words, the symmetric design $\{(c, \frac{1}{2}), (-c, \frac{1}{2})\}$ maximizes $|I(\alpha, \beta)|$, where c is obtained by maximizing $c^2 e^{2c}/(1+e^c)^4$. The maximizing value of c is $c_D = 1.5434$. Hence the D-optimal design consists of the points x_{1D} and x_{2D}, with weights $\frac{1}{2}$ each, satisfying $\alpha + \beta x_{1D} = -c_D$ and $\alpha + \beta x_{2D} = c_D$.

2.2. A-*optimality*

In order to obtain the A-optimal design, we shall minimize $\mathrm{Var}(\hat{\alpha}) + \mathrm{Var}(\hat{\beta})$, where $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimators of α and β and the variance being computed is the asymptotic variance. From the expression for the information matrix given in (2.4), we get

$$\mathrm{Var}(\hat{\alpha}) + \mathrm{Var}(\hat{\beta}) = \sum_{i=1}^{m} \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}[1 + x_i^2]/|I(\alpha, \beta)|$$

$$= \sum_{i=1}^{m} \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}\left[1 + \frac{(a_i - \alpha)^2}{\beta^2}\right] \Big/ |I(\alpha, \beta)|, \quad (2.8)$$

using (2.3). We do not have a complete solution to the A-optimality problem. However, if we restrict attention to symmetric designs, i.e., designs that are symmetric in the a_i's, then the A-optimal design can be easily obtained by applying Lemma 1. Note that for a symmetric design, (2.8) simplifies to

$$\mathrm{Var}(\hat{\alpha}) + \mathrm{Var}(\hat{\beta}) = \left[\frac{1}{\beta^2} \sum_{i=1}^{m} \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}[\alpha^2 + \beta^2 + a_i^2]\right] \Big/ \left[\frac{1}{\beta^2} \sum_{i=1}^{m} \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2} \times \sum_{i=1}^{m} \xi_i a_i^2 \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right]$$

$$= \frac{\alpha^2 + \beta^2}{\sum_{i=1}^{m} \xi_i a_i^2 e^{-a_i}/(1+e^{-a_i})^2} + \frac{1}{\sum_{i=1}^{m} \xi_i e^{-a_i}/(1+e^{-a_i})^2}, \quad (2.9)$$

where we have also used the expression for $|I(\alpha, \beta)|$ in (2.6). Now, let c satisfy (2.1) and (2.2), where we are using the fact that $e^{-a_i}/(1+e^{-a_i})^2 = e^{a_i}/(1+e^{a_i})^2$. Then, in the class of symmetric designs,

$$\mathrm{Var}(\hat{\alpha}) + \mathrm{Var}(\hat{\beta}) \geqslant \frac{\alpha^2 + \beta^2}{c^2 e^c/(1+e^c)^2} + \frac{1}{e^c/(1+e^c)^2}. \quad (2.10)$$

In other words, in the class of symmetric designs, the A-optimal design is given by $\{(c, \frac{1}{2}), (-c, \frac{1}{2})\}$, where c minimizes

$$\frac{\alpha^2 + \beta^2}{c^2 e^c/(1+e^c)^2} + \frac{1}{e^c/(1+e^c)^2}. \quad (2.11)$$

Once initial estimates of α and β are available, the A-optimal choice of c, say $c_A^{(1)}$ can be numerically obtained, by minimizing the expression in (2.11). For various values of α and β, Table 1 gives the values of $c_A^{(1)}$, numerically obtained. In the table, $A_{\mathrm{opt}}^{(1)}$ denotes the minimum value of the expression in (2.11), which is also the minimum

Table 1
Values of $c = c_A^{(1)}$ that minimizes (2.11), and $c_1 = c_{1A}^{(1)}$, $c_2 = c_{2A}^{(1)}$ and $\xi_1 = \xi_{1A}^{(1)}$ that minimize (2.12). $A_{\text{opt}}^{(1)}$ and $A_{\text{opt}}^{*(1)}$ denote the minimum values of (2.11) and (2.12), respectively

(α, β)	$c_A^{(1)}$	$A_{\text{opt}}^{(1)}$	$c_{1A}^{(1)}$	$c_{2A}^{(1)}$	$\xi_{1A}^{(1)}$	$A_{\text{opt}}^{*(1)}$	Loss of efficiency (%)
(10, 5)	2.3300	297.3141	2.3832	−2.3832	0.4056	287.2913	3.4887
(5, 5)	2.2464	126.0928	2.3065	−2.3065	0.3908	120.4794	4.6591
(1, 5)	2.1477	70.8927	2.1526	−2.1526	0.4647	70.5414	0.4979
(10, 2)	2.3175	249.4336	2.3954	−2.3954	0.3851	237.3101	5.1087
(5, 2)	2.1667	77.8308	2.3403	−2.3403	0.3043	68.1277	14.2425
(1, 2)	1.7550	20.8724	1.7701	−1.7701	0.3854	19.8340	5.2353
(10, 0.5)	2.3148	240.8808	2.3990	−2.3990	0.3804	228.2756	5.5219
(5, 0.5)	2.1424	69.1552	2.3932	−2.3932	0.2637	57.6540	19.9485
(1, 0.5)	1.3612	10.3111	1.2747	−1.2747	0.1968	7.5763	36.0972

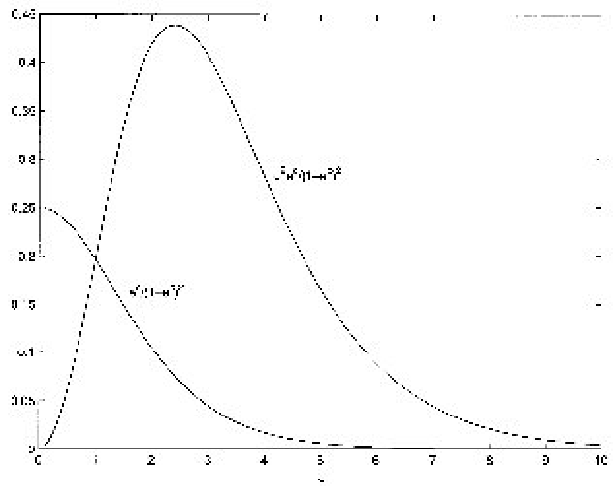

Fig. 1.

value of $\text{Var}(\hat{\alpha}) + \text{Var}(\hat{\beta})$ in the class of symmetric designs. The other quantities in Table 1 will be explained shortly.

In Fig. 1, we have plotted the functions $e^c/(1+e^c)^2$ and $c^2e^c/(1+e^c)^2$, for $c \geqslant 0$. The function $e^c/(1+e^c)^2$ is a strictly decreasing function of c, for $c \geqslant 0$, with maximum value of $\frac{1}{4}$ attained at $c = 0$. For $c \geqslant 0$, the function $c^2e^c/(1+e^c)^2$ increases, reaches a maximum at $c = 2.399$ (approximately), and then decreases. Consequently, if we are minimizing any decreasing function of $e^c/(1+e^c)^2$ and $c^2e^c/(1+e^c)^2$, the minimum will be at a value of c that will not exceed 2.399. Hence, the values of c_D and $c_A^{(1)}$ mentioned above, also do not exceed 2.399.

Even though we have derived the A-optimal design within the class of symmetric designs, it should be clear from the second expression in (2.8) that an A-optimal design within the class of all designs may not be a symmetric design, unless $\alpha = 0$. It appears difficult to characterize the A-optimal design within the class of all designs. We shall now restrict attention to a two point design $\{(c_1, \xi_1), (c_2, \xi_2)\}$ and numerically obtain the A-optimal design within the class of all such designs. Restricting attention to such two point designs, the A-optimal design problem reduces to that of minimizing

$$\frac{\xi_1 C_1[\beta^2 + (c_1 - \alpha)^2] + \xi_2 C_2[\beta^2 + (c_2 - \alpha)^2]}{(\xi_1 C_1 + \xi_2 C_2)(\xi_1 c_1^2 C_1 + \xi_2 c_2^2 C_2) - (\xi_1 c_1 C_1 + \xi_2 c_2 C_2)^2}, \tag{2.12}$$

where

$$C_i = \frac{e^{-c_i}}{(1 + e^{-c_i})^2}, \quad i = 1, 2. \tag{2.13}$$

The minimization of the expression in (2.12) can be done numerically with respect to c_1, c_2 and ξ_1 ($\xi_2 = 1 - \xi_1$). For various values of α and β, Table 1 also gives the values of c_1, c_2 and ξ_1, denoted by $c_{1A}^{(1)}$, $c_{2A}^{(1)}$ and $\xi_{1A}^{(1)}$ that minimizes (2.12). $A_{\text{opt}}^{*(1)}$ in Table 1 denotes the minimum value of (2.12). Also included are the percentage loss of efficiency of the symmetric A-optimal design $\{(c_A^{(1)}, \frac{1}{2}), (-c_A^{(1)}, \frac{1}{2})\}$, relative to the A-optimal design $\{(c_{1A}^{(1)}, \xi_{1A}^{(1)}), (c_{2A}^{(1)}, 1 - \xi_{1A}^{(1)})\}$. The expression for the percentage loss of efficiency is $(A_{\text{opt}}^{(1)}/A_{\text{opt}}^{*(1)} - 1) \times 100$. It is clear from the results in Table 1 that the loss of efficiency increases as β becomes smaller. Another interesting feature is that $c_{2A}^{(1)} = -c_{1A}^{(1)}$. That is, the A-optimal design is point symmetric. However, it is not weight symmetric, i.e., $\xi_{1A}^{(1)} \neq 0.5$, unless $\alpha = 0$. We have not been able to theoretically establish the point symmetry of the A-optimal design.

Our numerical results show that when α is replaced by $-\alpha$, the values of $c_{1A}^{(1)}$, $c_{2A}^{(1)}$ and $A_{\text{opt}}^{*(1)}$ do not change. However, $\xi_{1A}^{(1)}$ gets replaced by $1 - \xi_{1A}^{(1)}$. Thus, in the above table, we have reported numerical results only for a positive α.

3. Estimation of $\theta_1 = \alpha/\beta$ and β

The information matrix is now given by

$$I(\theta_1, \beta) = \begin{pmatrix} \beta^2 \sum_{i=1}^m \xi_i \dfrac{e^{-a_i}}{(1 + e^{-a_i})^2} & \sum_{i=1}^m \xi_i a_i \dfrac{e^{-a_i}}{(1 + e^{-a_i})^2} \\ \sum_{i=1}^m \xi_i a_i \dfrac{e^{-a_i}}{(1 + e^{-a_i})^2} & \dfrac{1}{\beta^2} \sum_{i=1}^m \xi_i a_i^2 \dfrac{e^{-a_i}}{(1 + e^{-a_i})^2} \end{pmatrix}, \tag{3.1}$$

where

$$a_i = \alpha + \beta x_i = \beta(x_i + \theta_1). \tag{3.2}$$

We shall first prove the following lemma. The lemma provides a result on weak supermajorization involving the vector of eigenvalues of $I(\theta_1, \beta)$ (see Marshall and Olkin (1979, p. 10) for the definition of weak supermajorization).

Lemma 2. *Let $I(\theta_1, \beta)$ be as given in* (3.1) *and let $I_c(\theta_1, \beta)$ denote the information matrix of the design $\{(c, \frac{1}{2}), (-c, \frac{1}{2})\}$, where c satisfies* (2.1) *and* (2.2). *Then the vector* *of eigenvalues of* $I_c(\theta_1,\beta)$ *is weakly supermajorized by the vector of eigenvalues of* $I(\theta_1,\beta)$.

Remark 1. Let $\boldsymbol{\lambda}(\theta_1,\beta)=(\lambda_1(\theta_1,\beta),\lambda_2(\theta_1,\beta))$ and $\boldsymbol{\lambda}_c(\theta_1,\beta)=(\lambda_{1c}(\theta_1,\beta),\lambda_{2c}(\theta_1,\beta))$ denote the vectors consisting of the eigenvalues of $I(\theta_1,\beta)$ and $I_c(\theta_1,\beta)$, respectively, where, we also assume $\lambda_1(\theta_1,\beta)\geqslant\lambda_2(\theta_1,\beta)$ and $\lambda_{1c}(\theta_1,\beta)\geqslant\lambda_{2c}(\theta_1,\beta)$. If $\phi(\boldsymbol{\lambda}(\theta_1,\beta))$ is a nonincreasing Schur-convex function of $\boldsymbol{\lambda}(\theta_1,\beta)$, then Lemma 2 implies that $\boldsymbol{\lambda}_c(\theta_1,\beta)$ minimizes $\phi(\boldsymbol{\lambda}(\theta_1,\beta))$ where c satisfies (2.1) and (2.2). Hence, if c_ϕ provides the minimum of $\phi(\boldsymbol{\lambda}_c(\theta_1,\beta))$ with respect to c, then $\{(c_\phi,\frac{1}{2}),(-c_\phi,\frac{1}{2})\}$ is an optimal design with respect to the optimality criterion ϕ. For example, the D-, A- and E-optimal design can be obtained by minimizing $|I_c(\theta_1,\beta)|^{-1}$, $\mathrm{tr}([I_c(\theta_1,\beta)]^{-1})$ and the maximum eigenvalue of $I_c(\theta_1,\beta)^{-1}$, respectively. Also note that $I_c(\theta_1,\beta)$ is a diagonal matrix with diagonal elements $\beta^2\mathrm{e}^{-c}/(1+\mathrm{e}^{-c})^2$ and $(1/\beta^2)c^2\mathrm{e}^{-c}/(1+\mathrm{e}^{-c})^2$. Hence, the vector $\boldsymbol{\lambda}_c(\theta_1,\beta)$ consists of the quantities $\beta^2\mathrm{e}^{-c}/(1+\mathrm{e}^{-c})^2$ and $(1/\beta^2)c^2\mathrm{e}^{-c}/(1+\mathrm{e}^{-c})^2$, ordered from the larger to the smaller. The approach to optimality via weak supermajorization is described in Bondar (1983), in a very general setup. In the context of binary data, Khan and Yazdi (1988) have also used majorization in order to derive D-optimal designs.

Proof of Lemma 2. It is well known that the vector of eigenvalues of a real symmetric matrix majorizes its vector of diagonal elements. In other words, the vector of diagonal elements of $I(\theta_1,\beta)$ is weakly supermajorized by the vector of eigenvalues of $I(\theta_1,\beta)$. When c satisfies (2.1) and (2.2), it follows that the vector of eigenvalues of $I_c(\theta_1,\beta)$, namely the vector

$$\left(\beta^2\frac{\mathrm{e}^{-c}}{(1+\mathrm{e}^{-c})^2},\frac{1}{\beta^2}c^2\frac{\mathrm{e}^{-c}}{(1+\mathrm{e}^{-c})^2}\right)'$$

is weakly supermajorized by the vector

$$\left(\beta^2\sum_{i=1}^{m}\xi_i\frac{\mathrm{e}^{-a_i}}{(1+\mathrm{e}^{-a_i})^2},\frac{1}{\beta^2}\sum_{i=1}^{m}\xi_i a_i^2\frac{\mathrm{e}^{-a_i}}{(1+\mathrm{e}^{-a_i})^2}\right)'.$$

Since the latter vector is the vector of diagonal elements of $I(\theta_1,\beta)$, the proof of Lemma 2 is complete.

As already pointed out, the D-optimal design derived for the estimation of α and β continues to be D-optimal for the estimation of θ_1 and β. We shall now consider A-optimality and E-optimality.

3.1. A-*optimality*

The A-optimal design minimizes

$$\frac{(1+\mathrm{e}^{-c})^2}{\mathrm{e}^{-c}}\left[\frac{1}{\beta^2}+\frac{\beta^2}{c^2}\right]. \tag{3.3}$$

Table 2
Values of $c = c_A^{(2)}$ that minimizes (3.3). $A_{\text{opt}}^{(2)}$ denotes the minimum value of (3.3)

β	$c_A^{(2)}$	$A_{\text{opt}}^{(2)}$
0.5	0.6925	20.3415
2	2.0510	11.8939
5	2.3843	57.4389

Table 3
Values of $c = c_E^{(2)}$ that minimizes (3.4). $E_{\text{opt}}^{(2)}$ denotes the minimum value of (3.4)

β	$c_E^{(2)}$	$E_{\text{opt}}^{(2)}$
0.5	0.2500	16.2513
2	2.3994	9.1069
5	2.3994	56.9179

The numerical minimization of (3.3) is easily accomplished, once a value of β is available. For a few values of β, Table 2 gives the optimum value of c, say $c_A^{(2)}$, that minimizes (3.3), along with the minimum value of (3.3), denoted by $A_{\text{opt}}^{(2)}$.

Sitter and Wu (1993) have also addressed the optimal design problem in the context of estimating θ_1 and β. However, the A-optimal design that they have constructed is for the estimation of θ_1 and $1/\beta$ (see Sitter and Wu, 1993, p. 331). In this case, the expression to be minimized becomes

$$\frac{(1+\mathrm{e}^{-c})^2}{\mathrm{e}^{-c}}\left[1+\frac{1}{c^2}\right].$$

3.2. E-*optimality*

The problem now is the minimization of

$$\max\left[\frac{(1+\mathrm{e}^{-c})^2}{\beta^2\mathrm{e}^{-c}}, \frac{\beta^2(1+\mathrm{e}^{-c})^2}{c^2\mathrm{e}^{-c}}\right]. \tag{3.4}$$

For various values of β, Table 3 gives the resulting optimum value of c, denoted by $c_E^{(2)}$, along with the minimum value of (3.4), denoted by $E_{\text{opt}}^{(2)}$.

4. Estimation of β and the 100γth percentile of $P(x)$

Let δ denote the 100γth percentile of $P(x)$. Then

$$\delta = \frac{l-\alpha}{\beta} \quad \text{where } l = \ln\left(\frac{100\gamma}{100(1-\gamma)}\right). \tag{4.1}$$

The information matrix of δ and β is given by

$$I(\delta,\beta)=\begin{pmatrix} \beta^2\sum_{i=1}^m \xi_i \dfrac{e^{-a_i}}{(1+e^{-a_i})^2} & -\sum_{i=1}^m \xi_i(a_i-l)\dfrac{e^{-a_i}}{(1+e^{-a_i})^2} \\ -\sum_{i=1}^m \xi_i(a_i-l)\dfrac{e^{-a_i}}{(1+e^{-a_i})^2} & \dfrac{1}{\beta^2}\sum_{i=1}^m \xi_i(a_i-l)^2\dfrac{e^{-a_i}}{(1+e^{-a_i})^2} \end{pmatrix}, \tag{4.2}$$

where

$$a_i=\alpha+\beta x_i=l+\beta(x_i-\delta). \tag{4.3}$$

It is easily seen that

$$|I(\alpha,\beta)|=\left[\sum_{i=1}^m \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right]\left[\sum_{i=1}^m \xi_i a_i^2 \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right]-\left[\sum_{i=1}^m \xi_i a_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right]^2. \tag{4.4}$$

The A-optimal design minimizes

$$\mathrm{Var}(\hat{\delta})+\mathrm{Var}(\hat{\beta})=\sum_{i=1}^m \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}\left[\beta^2+\frac{(a_i-l)^2}{\beta^2}\right]\Big/ |I(\delta,\beta)|, \tag{4.5}$$

where $\hat{\delta}$ and $\hat{\beta}$ denote maximum likelihood estimators and the variances under consideration are the asymptotic variances. If we consider only designs that are symmetric in the a_i's, then (4.5) simplifies to

$$\mathrm{Var}(\hat{\delta})+\mathrm{Var}(\hat{\beta})=\frac{1}{\beta^2}\left[\frac{\beta^4+l^2}{\sum_{i=1}^m \xi_i a_i^2 e^{-a_i}/(1+e^{-a_i})^2}+\frac{1}{\sum_{i=1}^m \xi_i e^{-a_i}/(1+e^{-a_i})^2}\right], \tag{4.6}$$

similar to (2.9). Now, we can apply Lemma 1 and conclude that the A-optimal design is given by $\{(c,\frac{1}{2}),(-c,\frac{1}{2})\}$, where c minimizes

$$\frac{1}{\beta^2}\left[\frac{\beta^4+l^2}{c^2e^c/(1+e^c)^2}+\frac{1}{e^c/(1+e^c)^2}\right]. \tag{4.7}$$

This is similar to the minimization of (2.11), once we have an initial value of β.

If we are interested in the 50th percentile of $P(x)$, then $l=0$ in (4.1). The 50th percentile is the quantity $-\alpha/\beta$. In other words, the estimation of the 50th percentile and β is equivalent to the estimation of α/β and β, the problem considered in Section 3.

5. Estimation of two percentiles of $P(x)$

Let δ_1 and δ_2, respectively, denote the $100\gamma_1$th and $100\gamma_2$th percentiles of $P(x)$, where we assume that $\gamma_1>\gamma_2$. Define

$$l_1=\ln\left(\frac{100\gamma_1}{100(1-\gamma_1)}\right) \quad \text{and} \quad l_2=\ln\left(\frac{100\gamma_2}{100(1-\gamma_2)}\right). \tag{5.1}$$

Then

$$\delta_1 = \frac{l_1 - \alpha}{\beta} \quad \text{and} \quad \delta_2 = \frac{l_2 - \alpha}{\beta}. \tag{5.2}$$

The information matrix of (δ_1, δ_2) is given by

$$I(\delta_1, \delta_2) = \frac{1}{(\delta_1 - \delta_2)^2} \times \begin{pmatrix} \sum_{i=1}^m \xi_i (a_i - l_2)^2 \frac{e^{-a_i}}{(1+e^{-a_i})^2} & -\sum_{i=1}^m \xi_i (a_i - l_1)(a_i - l_2) \frac{e^{-a_i}}{(1+e^{-a_i})^2} \\ -\sum_{i=1}^m \xi_i (a_i - l_1)(a_i - l_2) \frac{e^{-a_i}}{(1+e^{-a_i})^2} & \sum_{i=1}^m \xi_i (a_i - l_1)^2 \frac{e^{-a_i}}{(1+e^{-a_i})^2} \end{pmatrix}, \tag{5.3}$$

where

$$a_i = \alpha + \beta x_i = \frac{1}{\delta_1 - \delta_2}[(\delta_1 l_2 - \delta_2 l_1) + (l_1 - l_2)x_i]. \tag{5.4}$$

Then

$$\frac{(\delta_1 - \delta_2)^4}{(l_1 - l_2)^2} |I(\delta_1, \delta_2)| = \left[\sum_{i=1}^m \xi_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right] \left[\sum_{i=1}^m \xi_i a_i^2 \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right] - \left[\sum_{i=1}^m \xi_i a_i \frac{e^{-a_i}}{(1+e^{-a_i})^2}\right]^2. \tag{5.5}$$

The A-optimal design minimizes

$$\frac{\sum_{i=1}^m \xi_i (e^{-a_i}/(1+e^{-a_i})^2)[(a_i - l_1)^2 + (a_i - l_2)^2]}{\left[\sum_{i=1}^m \xi_i e^{-a_i}/(1+e^{-a_i})^2\right]\left[\sum_{i=1}^m \xi_i a_i^2 e^{-a_i}/(1+e^{-a_i})^2\right] - \left[\sum_{i=1}^m \xi_i a_i e^{-a_i}/(1+e^{-a_i})^2\right]^2}. \tag{5.6}$$

Once again, if we consider only designs that are symmetric in the a_i's, (5.6) simplifies to

$$\frac{2}{\sum_{i=1}^m \xi_i e^{-a_i}/(1+e^{-a_i})^2} + \frac{l_1^2 + l_2^2}{\sum_{i=1}^m \xi_i a_i^2 e^{-a_i}/(1+e^{-a_i})^2}.$$

Applying Lemma 1, we see that the A-optimal design is given by $\{(c, \frac{1}{2}), (-c, \frac{1}{2})\}$, where c minimizes

$$\frac{2}{e^c/(1+e^c)^2} + \frac{l_1^2 + l_2^2}{c^2 e^c/(1+e^c)^2}. \tag{5.7}$$

6. Concluding remarks

For a variety of estimation problems involving two parameters that are functions of α and β in (1.1), we have provided a unified approach for deriving D- and A-optimal designs by applying Lemma 1. In some cases, we have succeeded in deriving the A-optimal design only in the class of symmetric designs. No such restriction is needed for D-optimality. In one situation, we have also characterized the E-optimal design. It should be noted that restricting to the class of symmetric designs could result in considerable loss of efficiency with respect to the A-optimality criterion (see the numerical results in Table 1). All our optimal designs are two point designs. Numerical results show that the A-optimal design is likely to be point symmetric, but is not a symmetric design. Whether the A-optimal design is always a two point design is an open question.

Appendix Proof of Lemma 1

Note that we can assume without loss of generality that $a_i \geqslant 0$, since $e^{a_i}/(1+e^{a_i})^2 = e^{-a_i}/(1+e^{-a_i})^2$. We shall first prove the lemma for $m=2$. For $m=2$, we need to prove the following. Let p and q be nonnegative real numbers satisfying $p+q=1$. For a, $z \geqslant 0$, we shall show the existence of $c \geqslant 0$ satisfying

$$p\frac{e^a}{(1+e^a)^2} + q\frac{e^z}{(1+e^z)^2} = \frac{e^c}{(1+e^c)^2}, \tag{A.1}$$

$$pa^2\frac{e^a}{(1+e^a)^2} + qz^2\frac{e^z}{(1+e^z)^2} \leqslant c^2\frac{e^c}{(1+e^c)^2}. \tag{A.2}$$

We shall assume without loss of generality that $a < z$. We shall fix a and treat the left-hand side of (A.1) as a function of z. Thus, let

$$A(z) = p\frac{e^a}{(1+e^a)^2} + q\frac{e^z}{(1+e^z)^2}. \tag{A.3}$$

Since for any x, $0 < e^x/(1+e^x)^2 \leqslant \frac{1}{4}$, we also have $0 < A(z) \leqslant \frac{1}{4}$. Writing $w = e^c$, (A.1) is equivalent to $w/(1+w)^2 = A(z)$, which can be solved for $w > 1$. The solution is

$$w = \frac{(1-2A(z)) + \sqrt{1-4A(z)}}{2A(z)}.$$

(Recall that $0 < A(z) \leqslant \frac{1}{4}$.) Hence, c satisfying (A.1) is given by

$$c = c(z) = \ln w = \ln\left[\frac{(1-2A(z)) + \sqrt{1-4A(z)}}{2A(z)}\right]. \tag{A.4}$$

Since $e^z/(1+e^z)^2$ is a decreasing function of z, it follows from (A.2) that

$$a \leqslant c(z) \leqslant z. \tag{A.5}$$

We need to show that $c(z)$ satisfies (A.2), for $z \geqslant a$. Let

$$g(z) = z^2\frac{e^z}{(1+e^z)^2} \tag{A.6}$$

and let

$$f(z) = \{c(z)\}^2 \frac{e^{c(z)}}{(1+e^{c(z)})^2} - \left[pa^2 \frac{e^a}{(1+e^a)^2} + qz^2 \frac{e^z}{(1+e^z)^2} \right]$$
$$= g(c(z)) - [pg(a) + qg(z)]. \tag{A.7}$$

In order to show that $c(z)$ satisfies (A.2), we have to show that $f(z) \geqslant 0$ for $z \geqslant a$. From (A.4), it readily follows that $c(a) = a$ and hence $f(a) = 0$. Thus, in order to show that $f(z) \geqslant 0$ for $z \geqslant a$, it is enough to show that $f(z)$ is an increasing function of z. In other words, we have to show that $\mathrm{d}f(z)/\mathrm{d}z > 0$. From (A.7),

$$\frac{\mathrm{d}f(z)}{\mathrm{d}z} = \frac{\mathrm{d}g(c(z))}{\mathrm{d}z} - q\frac{\mathrm{d}g(z)}{\mathrm{d}z}$$
$$= \frac{\mathrm{d}g(c(z))}{\mathrm{d}c(z)} \frac{\mathrm{d}c(z)}{\mathrm{d}A(z)} \frac{\mathrm{d}(A(z))}{\mathrm{d}z} - q\frac{\mathrm{d}g(z)}{\mathrm{d}z}, \tag{A.8}$$

where $A(z)$ is given in (A.3). Using the definitions of $A(z)$, $c(z)$ and $g(z)$ given in (A.3), (A.4) and (A.6), respectively, straightforward calculation of the derivatives in the last expression in (A.8) gives

$$\frac{\mathrm{d}f(z)}{\mathrm{d}z} = \frac{c(z)}{e^{c(z)} - 1} [\{c(z)+2\} - e^{c(z)}\{c(z)-2\}] q \frac{e^z(e^z-1)}{(1+e^z)^3}$$
$$- q\frac{ze^z}{(1+e^z)^3} [(z+2) - e^z(z-2)].$$

Hence, $\mathrm{d}f(z)/\mathrm{d}z > 0$ if and only if

$$\frac{c(z)}{e^{c(z)} - 1} [\{c(z)+2\} - e^{c(z)}\{c(z)-2\}] \geqslant \frac{z}{e^z - 1} [(z+2) - e^z(z-2)]. \tag{A.9}$$

Since $c(z) \leqslant z$ (see (A.5)), (A.9) is established if we can show that the function

$$h(z) = \frac{z}{e^z - 1} [(z+2) - e^z(z-2)]$$

is a decreasing function of z. That is, we need to show that $\mathrm{d}h(z)/\mathrm{d}z < 0$. Now

$$\frac{\mathrm{d}h(z)}{\mathrm{d}z} = -2\frac{(z+1) + e^{2z}(z-1)}{(e^z-1)^2},$$

which can be shown to be less than zero for any $z > 0$. Thus, we have established (A.9) and hence the existence of $c \geqslant 0$ satisfying (A.1) and (A.2). In other words, we have established Lemma 1 for the case $m = 2$.

In order to prove the lemma for $m \geqslant 3$, i.e., in order to establish (2.1) and (2.2), let

$$\xi_{1*} = \xi_1 \Big/ \left(1 - \sum_{i=3}^{m} \xi_i\right) \quad \text{and} \quad \xi_{2*} = \xi_2 \Big/ \left(1 - \sum_{i=3}^{m} \xi_i\right). \tag{A.10}$$

Note that $\xi_{1*}+\xi_{2*}=1$. Proving the existence of c satisfying (2.1) and (2.2) is equivalent to showing that there exists c satisfying

$$\left(1-\sum_{i=3}^{m}\xi_i\right)\left[\xi_{1*}\frac{e^{a_1}}{(1+e^{a_1})^2}+\xi_{2*}\frac{e^{a_2}}{(1+e^{a_2})^2}\right]+\sum_{i=3}^{m}\xi_i\frac{e^{a_i}}{(1+e^{a_i})^2}=\frac{e^{c}}{(1+e^{c})^2},$$

$$\left(1-\sum_{i=3}^{m}\xi_i\right)\left[\xi_{1*}a_1^2\frac{e^{a_1}}{(1+e^{a_1})^2}+\xi_{2*}a_2^2\frac{e^{a_2}}{(1+e^{a_2})^2}\right]+\sum_{i=3}^{m}\xi_i a_i^2\frac{e^{a_i}}{(1+e^{a_i})^2}$$
$$\leqslant c^2\frac{e^{c}}{(1+e^{c})^2}. \tag{A.11}$$

Since we have already established the result for $m=2$, there exists c_* satisfying

$$\xi_{1*}\frac{e^{a_1}}{(1+e^{a_1})^2}+\xi_{2*}\frac{e^{a_2}}{(1+e^{a_2})^2}=\frac{e^{c_*}}{(1+e^{c_*})^2},$$

$$\xi_{1*}a_1^2\frac{e^{a_1}}{(1+e^{a_1})^2}+\xi_{2*}a_2^2\frac{e^{a_2}}{(1+e^{a_2})^2}\leqslant c_*^2\frac{e^{c_*}}{(1+e^{c_*})^2}. \tag{A.12}$$

In view of (A.11), (A.12) is established if we can show that there exists c satisfying

$$\left(1-\sum_{i=3}^{m}\xi_i\right)\frac{e^{c_*}}{(1+e^{c_*})^2}+\sum_{i=3}^{m}\xi_i\frac{e^{a_i}}{(1+e^{a_i})^2}=\frac{e^{c}}{(1+e^{c})^2},$$

$$\left(1-\sum_{i=3}^{m}\xi_i\right)c_*^2\frac{e^{c_*}}{(1+e^{c_*})^2}+\sum_{i=3}^{m}\xi_i a_i^2\frac{e^{a_i}}{(1+e^{a_i})^2}\leqslant c^2\frac{e^{c}}{(1+e^{c})^2}. \tag{A.13}$$

We note that the left-hand side expressions in (A.13) is similar to those in (2.1) and (2.2), except that (A.13) involves only $m-1$ terms. Proceeding as in the derivation of (A.13) given above, we can reduce the problem to the case of $m=2$. This completes the proof of Lemma 1.

References

Abdelbasit, K.M., Plackett, R.L., 1983. Experimental design for binary data. J. Amer. Statist. Assoc. 78, 90–98.

Bondar, J.V., 1983. Universal optimality of experimental designs: definitions and a criterion. Canad. J. Statist. 11, 325–331.

Ford, I., Torsney, B., 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.

Hedayat, A.S., Yan, B., Pezzuto, J.M., 1997. Modeling and identifying optimum designs for fitting dose-response curves based on raw optical density data. J. Amer. Statist. Assoc. 92, 1132–1140.

Khan, M.K., Yazdi, A.A., 1988. On D-optimal designs for binary data. J. Stat. Plann. Inference 18, 83–91.

Marshall, A.W., Olkin, I., 1979. Inequalities: Theory of Majorization and its Applications. Academic Press, New York.

Minkin, S., 1987. Optimal design for binary data. J. Amer. Stat. Assoc. 82, 1098–1103.

Pukelsheim, F., 1993. Optimal Design of Experiments. Wiley, New York.

Sitter, R.R., Wu, C.F.J., 1993. Optimal designs for binary response experiments: Fieller, D, and A criteria. Scand. J. Statist. 20, 329–342.

Wu, C.F.J., 1988. Optimal design for percentile estimation of a quantal response curve. In: Dodge, Y., Federov, V., Wynn, H.P. (Eds.), Optimal Design and Analysis of Experiments. Elsevier, Amsterdam, pp. 213–223.