Statistica Sinica **5**(1995), 405-419

OPTIMAL DESIGNS FOR BINARY RESPONSE EXPERIMENTS WITH TWO DESIGN VARIABLES

Randy R. Sitter and Ben Torsney

Carleton University and University of Glasgow

Abstract. D-optimal and c-optimal designs for binary response experiments with two design variables are considered. It is shown that with two design variables and a bounded design space, there exists a D-optimal design which has at most four or six symmetrically arranged support points, from which further reduction in the number of support points may sometimes be accomplished by using an asymmetric weighting. For c-optimality we identify when the support consists of 1, 2 or 3 points. The amount of numerical work required does not differ from the one design variable case.

Key words and phrases: Binary data, convex hull, logit, optimal design, probit.

1. Introduction

Optimal design for linear models has been extensively studied and is fairly simple since the information matrix is independent of the unknown regression parameters. In the case of generalized linear models (McCullagh and Nelder (1989)) less is known. Wu (1988) and Ford, Torsney and Wu (1992) (FTW) consider c-optimal and D-optimal designs for generalized linear models with simple linear effect and one design variable and discuss the interest in such optimal designs. In the case of the D criterion and binomial variation, they found that the equivalence theory of Kiefer and Wolfowitz (Fedorov (1972), Silvey (1980)) may not lead to a nice characterisation of optimal designs. By numerical investigation they found the D-optimal designs have three (instead of two) support points for some link functions. Torsney and Musrati (1993) continued this work. It is well established that if a D-optimal design has exactly k support points, where k is the number of parameters, then the optimal design weights are $1/k$. Otherwise numerical techniques are usually needed (Torsney and Alahmadi (1992) and Torsney (1983, 1988)). FTW also consider c-optimal designs in the same setting. Sitter and Wu (1993) use a different approach to obtain nice characterisations of D-, A- and F-optimal designs for binary response experiments with one design variable and explain some of the numerical results in FTW. Note that F-optimal designs minimise the length of a Fieller interval.

In this paper we study the more difficult situation of two design variables.

Though this has important applications in many areas, it has become a particularly important problem in the health protection areas. Humans are often exposed to more than one potentially hazardous substance simultaneously. For a recent overview see Krewski and Thomas (1992). In the context of linear models, the addition of more design variables does not significantly change the complexity of the problem or the nature of its solution, however in the context of binary response experiments this is not so. To see this, note that for typical binary response models, in the case of a single design variable there exists a unique D-optimal design for an unbounded design space. As we show in §3, in the case of two design variables the optimality criterion can be made arbitrarily large by choice of design. Thus the design problem is fundamentally different. A similar situation arises in linear models even for one design variable. Typically one considers a bounded design space, and chooses the optimal design within this space. Thus we consider some useful bounded design spaces for two design variables and obtain D- and c-optimal designs for some common dose-response models.

2. Formulation of Binary Response Design Problems

2.1. Binary response models

In a binary response experiment n_i subjects are administered two agents at dose levels $\mathrm{x}_i = (x_{1i}, x_{2i})^T$, $i = 1, \ldots, q$, where x_i represents a vector of two design variables selected from a design space $\mathcal{X} \subset \mathbb{R}^2$, which, as will be discussed later, must be bounded in some way. The outcome is binary, i.e., response or nonresponse, with probabilities $p(\mathrm{x}) = \mathrm{F}(\mathrm{x}; \theta)$ and $1 - p(\mathrm{x})$, respectively. Suppose the number of responses at dose level x_i is r_i, and $r_1, \ldots, r_q$ are independent binomial random variables, $r_i \sim \mathrm{Bin}(n_i, p(\mathrm{x}_i))$. Then the log-likelihood is

$$L(\theta) = \sum_{i=1}^{q} [r_i \log \mathrm{F}(\mathrm{x}_i; \theta) + (n_i - r_i) \log\{1 - \mathrm{F}(\mathrm{x}_i; \theta)\}] + C, \qquad (2.1)$$

where C is independent of θ. Throughout this article we will restrict attention to

$$\mathrm{F}(\mathrm{x}; \theta) = \Psi\{\beta^T(\mathrm{x} - \mu\mathbf{1})\}, \qquad (2.2)$$

where Ψ is a known monotone function with $\lim_{t\to-\infty} \Psi(t) = 0$, $\lim_{t\to\infty} \Psi(t) = 1$, with $\mu \in \mathbb{R}$, $\mathbf{1} = (1, 1)^T \in \mathbb{R}^2$ and $\beta \in \mathbb{R}^2$. This would be termed a simple additive risk model in the toxicology literature (Krewski, Colin and Dewanji (1990)). Krewski and Thomas (1992, p. 105) state, "[a]t lower doses, theoretical arguments suggest that risks may be near additive. Thus, additivity at low doses has been invoked as a working hypothesis by regulatory authorities in the absence of evidence to the contrary."

Let $\theta = (\mu, \beta_1, \beta_2)^T$. The Fisher information matrix under the model implied by (2.1) and (2.2), is

$$I(\theta) = n \begin{bmatrix} (\beta_1+\beta_2)^2 S_{00} & -\frac{\beta_1+\beta_2}{\beta_1} S_{01} & -\frac{\beta_1+\beta_2}{\beta_2} S_{02} \\ -\frac{\beta_1+\beta_2}{\beta_1} S_{01} & \frac{1}{\beta_1^2} S_{11} & \frac{1}{\beta_1\beta_2} S_{12} \\ -\frac{\beta_1+\beta_2}{\beta_2} S_{02} & \frac{1}{\beta_1\beta_2} S_{12} & \frac{1}{\beta_2^2} S_{22} \end{bmatrix}, \tag{2.3}$$

where $S_{00} = \sum \lambda_i \psi(z_i)$, $S_{0j} = \sum \lambda_i z_{ji} \psi(z_i)$, $S_{jj'} = \sum \lambda_i z_{ji} z_{j'i} \psi(z_i)$, $z_{ji} = \beta_j(x_{ji} - \mu)$, $z_i = z_{1i} + z_{2i}$, $\lambda_i = n_i/n$, $n = \sum n_i$, and $\psi(t) = [\Psi'(t)]^2/\{\Psi(t)[1 - \Psi(t)]\}$ for $j, j' = 1, 2$, and the summations are from 1 to q.

2.2. Design criteria

A good design will make $I(\theta)$ 'large' or $I^{-1}(\theta)$ 'small'. Two standard criteria are D-optimality and c-optimality. A c-optimal design minimises, for a given vector c, the approximate variance of $\mathrm{c}^T\hat{\theta}$, where $\hat{\theta}$ is the maximum likelihood estimator of θ, namely $\mathrm{c}^T I^{-1}(\theta)\mathrm{c}$. A D-optimal design maximizes $\det(I(\theta))$. This is attractive because under suitable assumptions $\det(I^{-1/2}(\theta))$ is proportional to the volume of the asymptotic confidence region for θ. Since these criteria depend on the unknown θ, we must assume that, for the purposes of design, a good initial estimate of θ is available. Robustness of optimal designs to a poor initial estimate of θ is often of practical concern. So a compromise design which has more dose levels but is near optimal, when benchmarked against the optimal design, is often used in practice.

We note that

$$\det(I(\theta)) = \frac{n^2\,(\beta_1+\beta_2)^2}{\beta_1^2\beta_2^2}\det(M),$$

where M is the matrix of $\{S_{ij}\}$ for $i, j = 0, 1, 2$ and can be written $M = \sum_i \lambda_i \mathrm{s}_i \mathrm{s}_i^T \psi(\mathbf{1}^T \mathrm{z}_i)$, with $\mathrm{s}_i^T = (1, \mathrm{z}_i^T)$. Also $\mathrm{c}^T I^{-1}(\theta)\mathrm{c} = \mathrm{c}_\beta^T M^{-1}\mathrm{c}_\beta$, where $\mathrm{c}_\beta = \Lambda\mathrm{c}$, with $\Lambda = \mathrm{diag}\{(\beta_1+\beta_2)^{-1}, -\beta_1, -\beta_2\}$. Thus a D-optimal design should maximise $\det(M)$ while a c-optimal design should minimise $\mathrm{c}_\beta^T M^{-1}\mathrm{c}_\beta$.

It is appropriate at this point to discuss the design space $\mathcal{X} \subset \mathbb{R}^2$. For a typical binary response model in the case of a single design variable, i.e. $z = \beta(x - \mu)$ with $x \in \mathcal{X} \subset \mathbb{R}$, there exist D- and c-optimal designs for an unbounded design space. Usually, however, a bounded design space is of particular interest for various practical reasons. For example: (1) the experimenter would usually like all the dose levels in the design to have a moderate response probability so as to avoid the problem of no responses or complete response at a particular dose level, since this yields no point estimate at that dose and contributes little information about the shape of the curve; (2) the experimenter may not wish to use too high a dose due to possible toxic side-effects of a drug; and (3) it may not be feasible or even possible to accurately produce low dose levels of a particular agent. In the one design variable case, consideration of these and perhaps other constraints would usually imply a bounded interval design space $\mathcal{X} = [x_{\min}, x_{\max}]$. This is due to the fact that the probability of response is monotonic in x for the most common models, so that an interval on x translates into an interval in the response probabilities. Of (1)–(3) above, (1) is almost always a consideration, however, in many cases the optimal design over the unbounded design space will satisfy (1), and thus may be of interest.

Interestingly, in the case of two design variables, if the design space is unbounded, the optimality criterion can be made arbitrarily large by choice of design. A similar situation arises in linear models even for one design variable. Thus, in the case of two design variables it is both desirable, for practical reasons, and necessary, for theoretical reasons, to consider a design space which is bounded in some way. Practical considerations such as (1)–(3) will guide the experimenter in the choice of bounded $\mathcal{X}$. Viewing (2.2), we see that (1) translates naturally into bounding $b_1 \leq z_1 + z_2 \leq b_2$, where $z_j = \beta_j(x_j - \mu)$ and b_1 and b_2 are experimenter chosen constants, since any dose combinations (x_1, x_2) which satisfy $z_1 + z_2 = constant$ have the same response probability. Consideration of (2) and (3) leads naturally to bounds like $b_1 \leq z_1 + z_2 \leq b_2$ or perhaps like $b_1 \leq z_1 \leq b_2$ and $b_3 \leq z_2 \leq b_4$, where the b_j are experimenter chosen constants.

It is clear, of course, that the experimenter's constraints can lead to a variety of shapes of bounded regions in $\mathbb{R}^2$. Though the principles for obtaining optimal designs remain the same for any of these, some will be easier to handle theoretically than others. We will restrict consideration in this paper to a class of regions which will be able to handle many practical situations and yet for which optimal designs are quite easy to obtain. They consist of regions which bound $z_1 + z_2$ and exactly one other linear combination of z_1 and z_2; that is they will be of the form $b_1 \leq z_1 + z_2 \leq b_2$ and $b_3 \leq a_1 z_1 + a_2 z_2 \leq b_4$ for some experimenter chosen constants a_1, a_2, and b_j; $j = 1, \ldots, 4$.

To obtain "exact" D-optimal or c-optimal designs we need to optimise our criteria over choices of q, of $\{\mathrm{x}_i\}$ and of $\{n_i\}$. Since the $\{n_i\}$ must be integers this is a difficult and often intractable problem. We instead consider the "continuous" setting in which n_i/n, and $n = \sum n_i$, are replaced by a real λ_i with $0 < \lambda_i < 1$ and $\sum \lambda_i = 1$.

2.3. A canonical problem

FTW propose a useful canonical form for the general k parameter situation, though they only use it for the two-parameter case. Their proposal involves a parameter dependent linear transformation of the design variables x. We have already resorted to such a transformation in the mapping from x_{ji} to z_{ji} but this was for notational convenience and does not precisely suit our present purposes. In effect we want to further transform the z_{ji}.

To derive this canonical form in the present context and establish notation, suppose that $\mathcal{Z} \subset \mathbb{R}^2$ is the image of $\mathcal{X}$ under the transformation $\mathrm{x} \rightarrow \mathrm{z}$ and let z be a typical element of $\mathcal{Z}$, and as before $\mathrm{s}^T = (1, \mathrm{z}^T)$. Consider the transformation $\mathrm{s} \rightarrow \mathrm{t} = B\mathrm{s}$ where B is a nonsingular 3×3 matrix which does not depend on the design variables and with its first and second rows, respectively, $(1,0,0)$ and $(0,1,1)$. Then $\mathrm{t}^T = (t_1, t_2, t_3)^T = (1, \mathrm{u}^T)$, where $\mathrm{u} = (u_1, u_2)^T \in \mathcal{U} \subset \mathbb{R}^2$ with $u_1 = t_2 = \mathbf{1}^T\mathrm{z}$, and $\mathcal{U}$ is the image of $\mathcal{Z}$ under this mapping. Further, since $\mathrm{s} = B^{-1}\mathrm{t}$, we have $M = (B^{-1})^T M_u B^{-1}$,

$$\det(M) = \det(M_u)/\left[\det(B)\right]^2 \quad \text{and} \quad \mathrm{c}_\beta^T M^1 \mathrm{c}_\beta = \mathrm{c}_\beta^T B^T M_u^{-1} B \mathrm{c}_\beta = \mathrm{c}_u^T M_u^{-1} \mathrm{c}_u, \tag{2.4}$$

where $\mathrm{c}_u = B\mathrm{c}_\beta$ and $M_u = \sum_i \lambda_i \mathrm{t}_i \mathrm{t}_i^T \psi(u_{1i})$ is the design matrix under a weighted linear model with weight function $\{\psi(u_1)\}^{-1/2}$ and design vector $\mathrm{u} \in \mathcal{U}$.

Thus a D-optimal design should maximise $\det(M_u)$ and a c-optimal design should minimise $\mathrm{c}_u^T M_u^{-1} \mathrm{c}_u$. These are respectively the D-optimal and c-optimal design problems for the above weighted linear model. It is these linear design problems which we classify as canonical design problems. They can be solved with optimal linear design problem tools.

Of course the transformations $\mathrm{x} \rightarrow \mathrm{z} \rightarrow \mathrm{u}$ depend on the parameter $\theta = (\mu, \beta_1, \beta_2)^T$. However, the dependence of optimal designs on the true value θ for given design space $\mathcal{X}$ is replaced in the transformed design problems by a design space $\mathcal{U}$ which varies with θ. Hence if we can solve the transformed design problem for arbitrary $\mathcal{U}$ we have implicitly solved the original design problems for arbitrary $\mathcal{X}$ and θ. What of the matrix B? Its partial definition only stipulates u_1, namely $u_1 = \beta^T(\mathrm{x} - \mu\mathbf{1})$. What should u_2 be?

If $\mathcal{X}$ is a bounded space then the choice does not matter and could be governed by aiming for simplicity of $\mathcal{U}$, or constraints which are natural to the application as discussed in the previous section. As we noted there, this paper will primarily consider constraints which bound $u_1 = z_1 + z_2$ and $u_2 = a_1 z_1 + a_2 z_2$ for some experimenter chosen constants a_1 and a_2, though many of the results will be useful more generally.

2.4. Geometrical characterizations of optimal supports

There are useful geometrical characterisations of D-optimal and c-optimal designs relative to an induced design space

$$G = \{\mathrm{g} \in \mathbb{R}^3 : \mathrm{g} = \{\psi(u_1)\}^{1/2}(1, u_1, u_2)^T, \ \ (u_1, u_2)^T \in \mathcal{U}\}.$$

D-optimal designs have as support points the points of contact between G and the smallest ellipsoid centered on the origin containing G (Silvey (1980, p. 41)). For c-optimal designs, letting $-G$ be the reflection of G about the origin, we must determine the convex hull of $\{G\} \cup \{-G\}$. The point where the vector c, extended if necessary, pierces this convex hull is a convex combination of some of the extreme points. These are the support points (see Elfving (1952), also Chernoff (1979)). Thus the design problem can be reduced to seeking a design on $\mathcal{U}$, and hence on G. For simplicity of notation, we will denote a design $\xi = \{(\lambda_i, \mathrm{u}_i);\ i = 1, \ldots, q\}$ if the design assigns weight λ_i to u_i for $i = 1, \ldots, q$.

At this point we need to be clear about the geometry of G. We suppose that g_1, and g_2 label axes in the horizontal plane and that g_3 labels the vertical axis. Consider the case $u_2 = 0$. This defines a horizontal cross-section of G which is identical to the corresponding induced design space for one explanatory variable. Let G_0 denote the corresponding slice of G, i.e.

$$G_0 = \{g \in \mathbb{R}^3 : g = \{\psi(u_1)\}^{1/2}(1, u_1, 0)^T, u_1 \in \mathbb{R}\}.$$

For the choices of $\Psi(\cdot)$ considered by FTW, by Sitter and Wu (1993) and by Torsney and Musrati (1993) this is a closed bounded trajectory anchored at the origin with u_1 potentially free to run from $-\infty$ to ∞. Five of these are given in Table 1, and Figure 1 gives G_0 and $-G_0$ with the convex hull also depicted for each of them.

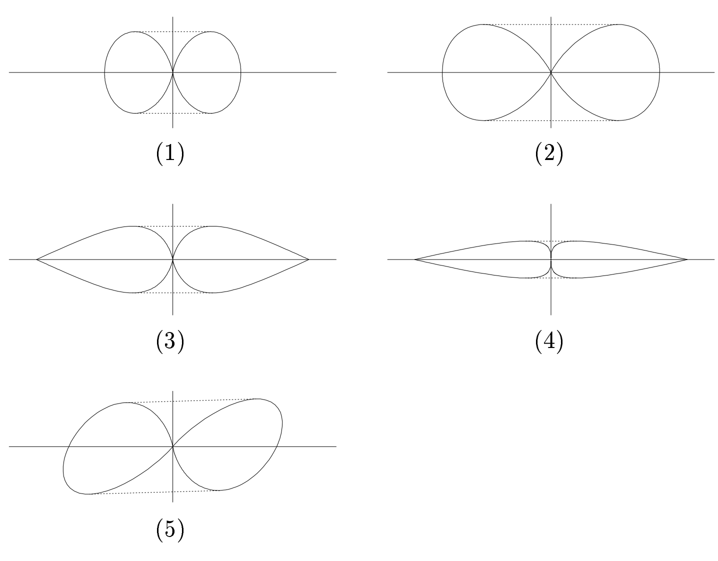

Figure 1. $\{G_0\} \cup \{-G_0\}$ for the five $\Psi(\cdot)$'s of Table 1. (1) Logit; (2) Probit; (3) Double exponential; (4) Double reciprocal; (5) Complementary log-log. The dotted lines complete the convex hull.

Table 1. $\underline{u}$ and $\overline{u}$ values for c-optimality and D-optimal designs for various $\Psi(\cdot)$

Case	Name	$\Psi(u)$	c-optimality		D-optimal design
			$\underline{u}$	$\overline{u}$	
1	Logit	$\{1+\exp(-u)\}^{-1}$	-2.4	2.4	$\{(.25, \pm 1.22, \pm 1)\}$
2	Probit	$\Phi(u)$	-1.58	1.58	$\{(.25, \pm .937, \pm 1)\}$
3	Double exponential	$\frac{(1+s)}{2} - \frac{s}{2}\exp(-\lvert u\rvert)$	-1.84	1.84	$\{(.094, \pm 1.59, \pm 1), (.312, 0, \pm 1)\}$
4	Double reciprocal	$\frac{(1+s)}{2} - \frac{s}{2}(1+\lvert u\rvert)^{-1}$	-1.62	1.62	$\{(.087, \pm 2^{1/2}, \pm 1), (.326, 0, \pm 1)\}$
5	Complementary log-log	$1-\exp(-\exp u)$	-2.07	1.27	$\{(.211, -1.08, \pm 1), (.289, .854, \pm 1)\}$

$^*s = \text{sign}(u)$.

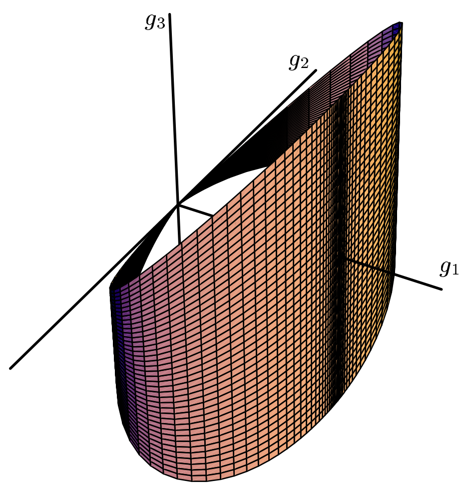

Figure 2. The induced design space, G_w, for two design variables and the logit of Table 1.

Now consider $u_1 = v_0$, a constant, so that $\text{g} = \{\psi(v_0)\}^{1/2}(1, v_0, u_2)^T$ defines a vertical line which will be unbound if u_2 is unbounded. Since G must be bounded, so must u_2. Without loss of generality, we assume that $-1 \le u_2 \le 1$. Hence for the widest possible range of u_1-values, the set G becomes

$$G_w = \left\{ \text{g} \in \mathbb{R}^3 : \text{g} = \{\psi(u_1)\}^{1/2}\ (1, u_1, u_2)^T, u_1 \in \mathbb{R}, -1 \le u_2 \le 1 \right\}.$$

This set resembles a, possibly asymmetric, vertically oriented signet ring beginning as a point at the origin ($u_1 = \pm\infty$) and linearly expanding as u_1 varies to a maximum length at the value of u_1 which maximizes $\psi(u_1)$. Figure 2 gives a three dimensional graphic of G_w for the logit of Table 1. A key feature of G_w is that it consists of a vertical surface bounded by ridges corresponding to $u_2 = \pm 1$. Each vertical line on the surface corresponds to a unique value of u_1.

We note that if the design space $\mathcal{X}$ is defined by additional constraints on x_1 and x_2, the corresponding induced design space G will be a subset of G_w. In particular, bounds on u_1 will correspond to the contiguous vertical section of G_w remaining after vertical cuts to G at the positions corresponding to these bounds. In the sequel, we discuss D- and c-optimal designs for only such constraints. However, once the above geometry is understood, obtaining optimal designs for more complicated bounds and constraints becomes much simpler.

3. D-Optimal Design

3.1. The case of $G = G_w$

We consider first the case $G = G_w$. It is immediately clear that any ellipsoid centered on the origin containing G can only touch G on the upper and lower ridges. Since the support points of the D-optimal design are the points of contact between G and the smallest such ellipsoid we conclude that D-optimal support points lie on these ridges and hence have $u_2 = \pm 1$. It is well established that, if there are k parameters, there exists a D-optimal design which has at least k and at most $k(k+1)/2$ support points. Since there are $k = 3$ parameters, there are at least 3 and at most 6 of them. Unfortunately, this would involve a high dimensional maximization over six possible u_1 values with their corresponding weights, as well as the various combinations of $u_2 = \pm 1$. However, the space G_w is symmetric in the vertical direction. This leads to the conjecture that D-optimal supports are such that if observations are taken at a particular value of u_1, then these are split equally between $u_2 = \pm 1$. This can be confirmed algebraically. That is, suppose that $\xi^{(1)} = \{(\lambda_i, \mathrm{u}_i);\ i = 1, \ldots, q\} = \{(\lambda_i, u_{1i}, u_{2i});\ i = 1, \ldots, q\}$ is any candidate design, i.e. $u_{2i} = \pm 1$, and let $\xi^{(2)} = \{(\lambda_i, u_{1i}, -u_{2i});\ i = 1, \ldots, q\}$, i.e. its reflection across $u_2 = 0$. If we let $\overline{\xi} = (\xi^{(1)} + \xi^{(2)})/2$, then the 02 and 12 components of $M_u(\bar{\xi})$ are 0, while the remaining components are the same as those of $M_u(\xi^{(1)})$ (in obvious notation similar to (2.3)). From this it is easy to show that $\det\{M_u(\overline{\xi})\} \geq \det\{M_u(\xi^{(1)})\}$, with strict inequality if the 02 and/or 12 components of M_u are not zero. This implies that for any such design there exists a symmetric design, with possibly more support points, which is better. So, in order to find a D-optimal design, we need only consider designs of the form

$$\xi = \{(\lambda_i/2, u_{1i}, 1), (\lambda_i/2, u_{1i}, -1),\ i = 1, \ldots, q^*\}\,. \tag{3.1}$$

This design may however not be unique and may not have minimum support among all designs which attain the maximum value of the determinant.

This essentially reduces the complexity of the problem to that of the one design variable situation. We must choose q^*, u_{1i}, and λ_i^* for $i = 1, \ldots, q^*$, with $\sum \lambda_i^* = 1$ and $\lambda_i^* > 0$, to maximise $D = S_{u00}(S_{u00}S_{u11} - S_{u01}^2)$, where S_{uij} denotes the ijth component of M_u. As a direct result of Caratheodory's Theorem (Silvey (1980, p. 72)), it follows that $q^* = 2$ or 3. Since each u_{1i} implies 2 support points, this implies a total support of 4 or 6 points. Furthermore, this reduces the numerical maximization to at most 5 variables; at most $q^* = 3$ u_{1i} values plus their corresponding λ_i^* weights, noting that $\lambda_3^* = 1 - \lambda_1^* - \lambda_2^*$.

It was shown in Sitter and Wu (1993) that, in the two parameter case, there may exist more than one design with differing sets of support points which attain the maximum determinant value. A similar property can exist in our case. For example, it may be that there exists an asymmetric design with positive weights on only a subset of the support points of the obtained symmetric design that has the same determinant value and smaller support.

If $\psi(\cdot)$ is symmetric about zero, then there exists a design which is symmetric about zero in u_1 in terms of support points and weights which is D-optimal, though the resulting design may not have minimum support. We can obtain it using methods similar to those of Sitter and Wu (1993), where they consider various conditions on $\psi(\cdot)$. In what they term cases (i) and (ii) (p. 332) we can find this design by considering only designs of the form

$$\xi = \left\{ \left(\frac{\lambda}{4}, \pm u_1, 1\right), \quad \left(\frac{\lambda}{4}, \pm u_1, -1\right), \quad \left(\frac{1-\lambda}{2}, 0, 1\right), \quad \left(\frac{1-\lambda}{2}, 0, -1\right) \right\},$$

choosing $u_1 \in \mathbb{R}$ and $\lambda \in (0, 1]$ to maximise D; a 2 variable maximization. Note that $\lambda = 1$ corresponds to a four point design. Note also that, if we let

$$h(z) = \frac{d}{dz}[z^2\psi(z)] / \frac{d}{dz}\psi(z) = zf(z),$$

where $f(z) = z + 2\psi(z)/\psi'(z)$, then $h'(z) > 0$ for $z > 0$ implies case (i) and $h''(z) > 0$ for $z > 0$ implies case (i) or (ii). It is not difficult to show that the $\Psi(\cdot)$ numbered 1 in Table 1 (logit) satisfies $h'(z) > 0$ for $z > 0$, while those numbered 3 and 4 satisfy $h''(z) > 0$ for $z > 0$. We have no rigorous proof that number 2 (probit) satisfies $h'(z) > 0$ for $z > 0$ though a plot of $h(\cdot)$ seems to indicate that it is so.

Approximate D-optimal designs on the transformed design space $\mathcal{U}$, obtained by numerically solving the above maximization problem, for five choices of $\Psi(\cdot)$ are listed in Table 1. We see that the logit and the probit, which have symmetric $\psi(\cdot)$, yield four-point designs symmetric in u_1, while the complementary log-log, which has an asymmetric $\psi(\cdot)$, yields a four-point design which is asymmetric in u_1. This is similar to the two parameter case, where the logit and probit yield two-point designs symmetric in u_1 and the complementary log-log yields an asymmetric design. In all three cases the optimal values of u_1 in the two- and three-parameter cases differ. We note this because the same is not true for the double exponential and the double reciprocal. In these two cases we see that the D-optimal designs have six points symmetric in u_1. Sitter and Wu (1993) show that in the two-parameter case these yield three-point designs symmetric in u_1 with exactly the same u_1 values as here. In fact, one can show that if $\psi(\cdot)$ falls into their case (ii) with a three point solution, which these two do, this will always hold. Though the proof involves a very easy argument within the geometrical framework of Sitter and Wu, we do not present it here to avoid a lengthy introduction of their geometry.

3.2. The case of $G \subset G_w$

We consider the case $u_1 \in [a,b]$, where $a < 0$ and $b > 0$, so that

$$G = \left\{ \mathrm{g} \in \mathbb{R}^3 : \mathrm{g} = \{\psi(u_1)\}^{1/2}(1,u_1,u_2)^T, a \le u_1 \le b, -1 \le u_2 \le 1 \right\}.$$

Again we can argue that support points can only be on the ridges of G and we can restrict attention to weights equally distributed between $u_2 = \pm 1$, and thus to the simplified designs considered for G_w with the proviso that the u_1-values must lie in $[a,b]$.

Results parallel those, empirical and analytical, for the one explanatory variable case. A major issue is the number of support points. Let a^{**}, b^{**} denote the support points (u_1-values) of D-optimal designs on G_w. Using geometrical arguments similar to Sitter and Wu (1993, p. 332-333) it is not difficult to show that for the logit and probit models of Table 1, $a^{**} = -b^{**}$, only two support points are needed, and

(a) If $[a,b] \ni a^{**}, b^{**}$ then G and G_w have the same D-optimal design.
(b) Suppose that $a > a^{**}$. Then the support points are a and $\min\{b, u_1^*\}$, where u_1^* maximizes D assuming an optimally weighted design of the form (3.1) with $q^* = 2$ and support points a and u_1^*.
(c) Suppose that $b < b^{**}$. Then the support points are b and $\max\{a, u_1^*\}$, where u_1^* maximizes D assuming an optimally weighted design of the form (3.1) with $q^* = 2$ and support points u_1^* and b.
(d) If $[a,b] \subset [a^{**}, b^{**}]$ the support points are a and b.

For the double exponential and double reciprocal a solution will be more complex. The number of support points can be 2 or 3, and in each case these can include both or only one or neither of the two endpoints a and b. Torsney and Musrati (1993) find results similar to the above for the one design variable case using numerical investigations.

To see how one can use these D-optimal designs on $\mathcal{U}$ for a particular application, suppose that the experimenter has initial estimates of (μ, β_1, β_2) of $(10, 1, 1)$ and believes that the logit will be an adequate model for the experiment. Practical constraints are considered and it is felt that with the number of subjects available, the design space should be constrained so that the probability of response is between 0.1 and 0.9 so as to avoid the situation of no or complete response at a particular dose. This implies the constraint $-2.2 \leq z_1 + z_2 \leq 2.2$. It is also felt that the doses of the two agents should not be too different. From this it is decided to constrain $-2 \leq z_1 - z_2 \leq 2$. So let $u_1 = z_1 + z_2$ and $u_2 = (z_1 - z_2)/2$, i.e. $-1 \leq u_2 \leq 1$. We see from Table 1 that this situation is that of (a) above, i.e. $[a, b] \ni a^{**}, b^{**}$, and thus the D-optimal design for G_w is optimal for this design space. The D-optimal design is the four points $(.25, 1.22, 1)$, $(.25, 1.22, -1)$, $(.25, -1.22, 1)$, $(.25, -1.22, -1)$, which translates back into four points in the original design space, (λ, x_1, x_2), $(.25, 11.61, 9.61)$, $(.25, 9.61, 11.61)$, $(.25, 10.39, 8.39)$, $(.25, 8.39, 10.39)$. Note that the response probabilities are 0.772, 0.772, 0.228 and 0.228.

4. c-Optimal Design

4.1. The case of $G = G_w$

We again consider the case $G = G_w$. As was mentioned in §2.4, the point where the vector c, extended if necessary, intersects the convex hull of $\{G\} \cup \{-G\}$ is a convex combination of extreme points and it is these extreme points which form the support of the c-optimal design. In addition, Fellman (1974, Theorem 3.1.4) proves that this c-optimal design can have at most k linearly independent points from G (recall k is the number of parameters). Once the support points, $g_1, \ldots, g_s$ are obtained, explicit c-optimal weights are available, namely $\lambda_j^* = |a_j| / \sum_{i=1}^{s} |a_i|$, where $\mathrm{a} = (a_1, \ldots, a_s)^T = (V^T V)^{-1} V^T \mathrm{c}$, and $V = (g_1, \ldots, g_s)$ (Pukelsheim and Torsney (1991) or Kitsos, Titterington and Torsney (1988)). It is also the case that these weights are the convex weights of the above convex combination of extreme points which is proportional to c. Thus, to completely determine a c-optimal design we need only obtain these support points.

To do this, we need to identify the boundary of the convex hull of $G \cup \{-G\}$. It is helpful to consider first the case $u_2 = 0$, which corresponds to the "single explanatory variable" case. FTW describe the boundary of the convex hull of $G_0 \cup \{-G_0\}$ given in Figure 1. It consists of the arc of G_0 from A to B, its image about the origin, and the parallel lines joining A to $-B$ and B to $-A$, where: (i) $A = \{\psi(\underline{u})\}^{1/2}(1, \underline{u}, 0)^T$, $B = \{\psi(\overline{u})\}^{1/2}(1, \overline{u}, 0)^T$, and $\underline{u}$, $\overline{u}$ are the solutions to

$$r(\underline{u}) = s(\overline{u}, \underline{u}) \quad \text{and} \quad r(\overline{u}) = s(\overline{u}, \underline{u}) \quad \text{given} \quad \underline{u} \text{ and } \overline{u};$$

and (ii) $r(z)$ is the slope of the curve $\{\psi(z)\}^{1/2}(1,z)^T$ at z, namely

$$r(z)=\frac{d}{dz}\{w(z)z\}/\frac{d}{dz}\{w(z)\}=z+w(z)/w'(z),$$

where $w(z)=\{\psi(z)\}^{1/2}$, while $s(z_1,z_2)=\{w(z_1)z_1+w(z_2)z_2\}/\{w(z_1)+w(z_2)\}$. FTW determine $\underline{u}$ and $\overline{u}$ for nine choices of $\psi(\cdot)$ including the five considered in Table 1. These same values will be the critical points in our context, and thus we need do *no extra numerical work*.

Given this, construction of the boundary of the convex hull of $G\cup\{-G\}$ becomes clear. It is like a, possibly asymmetric, train carriage with: a curved front and back, reflections of each other; a curved roof and curved floor, again reflections of each other; and two parallel planar sides. Figure 1 would represent top-views of this train carriage with the two parallel dotted lines being the sides.

The 'front' consists of that section of G, see Figure 2, between $\underline{u}$ and $\overline{u}$; the same quantities as those given by FTW in their Table 4 are reproduced for our five examples in Table 1. The 'back' is the reflection of this about the origin. One planar side is a quadrilateral with vertices $[\{\psi(\underline{u})\}^{1/2}(1,\underline{u},\pm1)^T$, $-\{\psi(\overline{u})\}^{1/2}(1,\overline{u},\pm1)^T]$, with the other having vertices $[-\{\psi(\underline{u})\}^{1/2}(1,\underline{u},\pm1)^T$, $\{\psi(\overline{u})\}^{1/2}$ $(1,\overline{u},\pm1)^T]$. The 'roof' can be thought of as a contiguous sequence of lines parallel to the side planes, namely the lines joining $[\{\psi(u)\}^{1/2}(1,u,1)^T]$ and $[-\{\psi(u)\}^{1/2}(1,u,-1)^T]$ for each $u\in[\underline{u},\overline{u}]$. Similarly the 'floor' comprises the lines joining $[\{\psi(u)\}^{1/2}(1,u,-1)^T]$ and $[-\{\psi(u)\}^{1/2}(1,u,1)^T]$. A final point of note is that the ridges of G between $\underline{u}$ and $\overline{u}$ and their reflections about the origin $[\pm\{\psi(u)\}^{1/2}(1,u,\pm1)^T,\quad u\in[\underline{u},\overline{u}]]$, i.e. the four ridges where the 'front' and 'back' meets the 'roof' and 'floor', respectively, form the extreme points of this boundary.

It is now easy to provide a descriptive solution to c-optimal designs. The solution depends on where the vector c_u, extended if necessary, pierces the boundary of $G\cup\{-G\}$. Denote this point by $\mathrm{c}_b=(c_{1b},c_{2b},c_{3b})^T$, so that $\mathrm{c}_u\propto\mathrm{c}_b$. We distinguish various cases. Support points are extreme points of the boundary of the convex hull of $G\cup\{-G\}$.

(i) c_b lies on one of the above four ridges, i.e. $\mathrm{c}_b=\pm\{\psi(\tilde{u})\}^{1/2}(1,\tilde{u},\pm1)^T$ for some $\tilde{u}\in[\underline{u},\overline{u}]$. Then the c-optimal design is the one-point design with all the weight at $\tilde{u}$.
(ii) c_b lies on the 'front' or 'back' surfaces, i.e. $\mathrm{c}_b=\pm\{\psi(\tilde{u})\}^{1/2}(1,\tilde{u},u_2)^T$, $\tilde{u}\in[\underline{u},\overline{u}]$, $-1<u_2<1$. Then c_b is a convex combination of $[\{\psi(\tilde{u})\}^{1/2}(1,\tilde{u},\pm1)^T]$ or of $[-\{\psi(\tilde{u})\}^{1/2}(1,\tilde{u},\pm1)^T]$ and the c-optimal design has two support points, namely $(\tilde{u},-1)$ and $(\tilde{u},1)$.
(iii) c_b lies on the 'roof' or 'floor'. Then for some $\tilde{u}\in[\underline{u},\overline{u}]$, c_b lies on the line joining $[\{\psi(\tilde{u})\}^{1/2}(1,\tilde{u},1)^T]$ to $[-\{\psi(\tilde{u})\}^{1/2}(1,\tilde{u},-1)^T]$ ('roof') or on the line

joining $[\{\psi(\tilde{u})\}^{1/2}(1,\tilde{u},-1)^T]$ to $[-\{\psi(\tilde{u})\}^{1/2}(1,\tilde{u},1)^T\}$ ('floor'). The c-optimal design has two support points, namely $(\tilde{u},\pm1)$.
(iv) c_b lies on either of the two side planes (quadrilaterals). In this case c_b is typically equal to several convex combinations of the four vertices of the quadrilateral. In particular, it should be a convex combination of at least one set of three vertices. It may be a convex combination of only two of them if it lies on a diagonal and of course c_b may actually coincide with one of the vertices. The implication is that for a given c there are various possible c-optimal designs whose supports are subsets of the four points $(\underline{u},\pm1)$, $(\overline{u},\pm1)$. At least one of these designs has at most three points.

This completes the descriptive solution. Of course it would be possible to translate it into algebraic rules but we do not feel this would give further insight.

Note that it will always be undesirable to use a 1- or 2- point design even though it is c-optimal since it will not even allow estimation of all the parameters in the model. However, these designs may be useful as a benchmark or when applied sequentially. See Wu (1988), FTW, and Sitter and Wu (1993) for further related discussion.

4.2. The case of $G \subset G_w$

We now consider the case $u_1 \in [a,b]$ so that

$$G = \{g \in \mathbb{R}^3 : g = \{\psi(u_1)\}^{1/2}(1,u_1,u_2)^T,\ \ a \le u_1 \le b,\ \ -1 \le u_2 \le 1\}$$

and make the same assumptions about G_0 as in §3.2. The boundary of the convex hull of $G \cup \{-G\}$ then has an identical 'train carriage' form to that of $G = G_w$, but with potentially different $\underline{u}$ and $\overline{u}$. Their values are the same as for the one explanatory variable case of FTW. There are four cases to be distinguished depending on the relationship of a and b to the values of $\underline{u}$ and $\overline{u}$ for G_w. Denote these $\underline{u}$ and $\overline{u}$ values for G_w as a^{**} and b^{**}.
(a) $b \ge b^{**}$ and $a \le a^{**}$; then $\underline{u} = a^{**}$ and $\overline{u} = b^{**}$.
(b) $b \le b^{**}$ and $a \ge a^{**}$; then $\underline{u} = a$ and $\overline{u} = b$.
(c) $b < b^{**}$ and $a \le a^{**}$; then $\overline{u} = b$ and $\underline{u} = \max\{a, u_b\}$, where u_b solves $r(u) = s(b,u)$.
(d) $b \ge b^{**}$ and $a > a^{**}$; then $\underline{u} = a$ and $\overline{u} = \min\{b, u_a\}$, where u_a solves $r(u) = s(a,u)$.

Applying the results of §4.1 yields c-optimal designs.

5. Concluding Remarks

We have considered D- and c-optimal designs for binary response models with two design variables, and have given straight-forward methods for obtaining them, which do not differ in numerical complexity from, and allow the use of some of the completed numerical work of, the one variable case. We feel this work has also revealed some interesting directions for further research. We are presently considering extensions to generalized linear models, multiple design variables, and bounded design spaces of a more complex nature. Some recent results on multiple design variables for generalized linear models appear in Burridge and Sebastiani (1994), Sitter and Torsney (1992) and Cao (1992).

Acknowledgement

R. R. Sitter was supported by the Natural Sciences and Engineering Research Council of Canada.

References

Burridge, J. and Sebastiani, P. (1994). D-optimal designs for generalised linear models with variance proportional to the square of the mean. *Biometrika* **81**, 295-304.

Cao, P. (1992). On multifactor D-optimal designs for generalized linear models with poisson and binary error structures. Ph.D. Thesis, Department of Mathematics and Statistics, Queens University, Kingston, Ontario, Canada.

Chernoff, H. (1979). Sequential Analysis and Optimal Design. Society for Industrial and Applied Mathematics, Philadelphia.

Elfving, G. (1952). Optimum allocation in linear regression theory. Ann. Math. Statist. **23**, 255-262.

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

Fellman, J. (1974). On the allocation of linear observations. Commentationes Physico-Mathematicaa (Finska Vetepskaps Societaten) **44**, 27-77.

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.

Kitsos, C. P., Titterington, D. M. and Torsney, B. (1988). An optimal design problem in rhythmometry. Biometrics **44**, 657-671.

Krewski, D. and Thomas, R. D. (1992). Carcinogenic mixtures. Risk Anal. **12**, 105-113.

Krewski, D., Colin, D and Dewanji, A. (1990). Statistical considerations in the design of toxicological experiments with chemical mixtures. Progress in Predictive Toxicology (Edited by D. B. Clayson, I. C. Munro, P. Shubik and J. A. Swenberg), 335-348, Elsevier Science Publishers B.V. (Biomedical Division).

McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models, 2nd edition. Chapman and Hall, London.

Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. Ann. Statist. **19**, 1614-1625.

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

Sitter, R. R. and Torsney, B. (1992). D-optimal designs for generalized linear models. Technical Report No. 211, Technical Report Series of the Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada.

Sitter, R. R. and Wu, C. F. J. (1993). Optimal designs for binary response experiments: Fieller, D, and A criteria. Scand. J. Statist. **20**, 329-341.

Torsney, B. (1983). A moment inequality and monotonicity of an algorithm. In Proc., Internat. Symp. on Semi-infinite Programming and Appl. (Edited by K. O. Kortanek and A. V. Fiacco) at Univ. of Texas, Austin. Lecture Notes in Econom. and Math. Sys. **215**, 249-260.

Torsney, B. (1988). Computing optimizing distributions with applications in design, estimation and image processing. Optimal Design and Analysis of Experiments (Edited by Y. Dodge, V. Fedorov and H. P. Wynn), 361-370, Elsevier Science Publishers B. V., North Holland.

Torsney, B. and Musrati, A. K. (1993). On the construction of optimal designs with applications to binary response and to weighted regression models. Proc. 3rd Model Oriented Data Analysis Workshop in Petrodvorets, Russia (Edited by W. G. Muller, H. P. Wynn, A. A. Zhigljavsky), 37-52, Physica-Verlag.

Torsney, B. and Alahmadi, A. M. (1992). Further development of algorithms for constructing optimizing distributions. Model Oriented Data Analysis. Proc. 2nd IIASA Workshop in St. Kyrik, Bulgaria (Edited by V. Fedorov, W. G. Muller, I. N. Vuchkov), 121-129, Physica-Verlag.

Wu, C. F. J. (1988). Optimal design for percentile estimation of a quantal response curve. Optimal Design and Analysis of Experiments (Edited by Y. Dodge, V. Fedorov and H. P. Wynn), 213-223, Elsevier Science Publishers B. V., North Holland.

Department of Mathematics and Statistics, Carleton University, Ottawa K1S 5B6, Canada.

Department of Statistics, University of Glasgow, Mathematics Building, University Gardens, Glasgow G12 8QW, U.K.

(Received October 1993; accepted January 1995)