Statistica Sinica **7**(1997), 941-955

OPTIMAL TWO-STAGE DESIGNS FOR BINARY RESPONSE EXPERIMENTS

R. R. Sitter and B. E. Forbes

Simon Fraser University and Carleton University

Abstract: In typical binary response models, the information matrix depends both on the design and the unknown parameters of interest. Thus to obtain optimal designs, one must have 'good' initial parameter estimates. Often this is not the case. One solution which may be applicable in some settings is to perform the experiment in two (or more) stages, that is, use an initial design to get parameter estimates and then treating these as the true parameter values choose a second stage design so that the combined first and second stage design is optimal in some sense. In this article, we consider a class of symmetric binary response models which includes the common logit and probit models, and show that for any of the main optimality criteria in the literature (eg. A-, D-, E-, F-, G- and c-optimality) the optimal second stage design will consist of two points symmetrically placed about the ED50 (the 50% response dose) with possibly different weights at each point. We go on to give examples where one can further reduce the resulting optimization to a one variable maximization. In the process some insight is gained into how the second stage design corrects skewness in the first stage design.

Key words and phrases: Binary data, confidence interval, Fieller's theorem, logit, probit.

1. Introduction

In a binary response experiment n_i subjects are administered a stimulus at a dose level x_i, possibly on a transformed scale, for $i = 1, \ldots, k$. The outcome is binary, i.e., response or non-response, with probabilities $p(x) = \mathrm{F}(x;\ \underline{\theta})$ and $q(x) = 1 - p(x)$, respectively. Suppose the number of responses at dose level x_i is r_i, and $r_1, \ldots, r_k$ are independent binomial random variables, $r_i \sim \mathrm{Bin}(n_i, p(x_i))$. Then the log-likelihood is

$$l(\underline{\theta}) \propto \sum_{i=1}^{k} [r_i \log p(x_i) + (n_i - r_i) \log q(x_i)]. \tag{1.1}$$

In this article we will restrict attention to

$$\mathrm{F}(x \mid \underline{\theta}) = \Psi\{\beta(x - \mu)\}, \tag{1.2}$$

where Ψ is a known strictly monotone function with $\lim_{t\to-\infty} \Psi(t) = 0$, $\lim_{t\to\infty} \Psi(t) = 1$, and Ψ' symmetric about zero. The two common models fall into this class: the logit, $\Psi(t) = (1 + e^{-t})^{-1}$; and the probit, $\Psi(t) = \int_{-\infty}^{t} (2\pi)^{-1/2} e^{-u^2/2} du$.

The classical optimal design problem for these models has received attention in the literature. Within this framework, an "exact" optimal (one stage) design problem is to choose k, $\{x_i\}$ and $\{n_i\}$ to maximize some measure of the information about (μ, β) yielded by the data. This measure is usually a function $\phi(\cdot)$ applied to the Fisher information matrix for the maximum likelihood estimator, $\underline{\hat{\theta}} = (\hat{\mu}, \hat{\beta})^T$, which in this case is

$$nM(\xi, \underline{\theta}) = n \begin{pmatrix} \beta^2 S_0 & -S_1 \\ -S_1 & \frac{1}{\beta^2} S_2 \end{pmatrix},$$

where $S_0 = \sum_{i=1}^{k} \lambda_i \psi(z_i)$, $S_1 = \sum_{i=1}^{k} \lambda_i z_i \psi(z_i)$, $S_2 = \sum_{i=1}^{k} \lambda_i z_i^2 \psi(z_i)$, $z_i = \beta(x_i - \mu)$, $\lambda_i = n_i/n$, $n = \sum_i n_i$, $\psi(t) = [\Psi'(t)]^2/[\Psi(t)\{1 - \Psi(t)\}]$, and $\xi = \{(\lambda_i, z_i) : i = 1, \ldots, k\}$ denotes the design measure. Thus the measure of information, $\phi\{M(\xi, \underline{\theta})\}$, depends on the unknown values of μ and β, and good initial estimates are needed. For (1.1) and (1.2) this dependence enters through $z = \beta(x - \mu)$ and the problem reduces to finding optimal k and ξ.

Since this is a difficult and often intractable optimization problem, usually the "continuous" setting is instead considered. That is, instead of maximizing $\phi\{M(\xi, \underline{\theta})\}$ over all finite design measures of the type described above, we maximize over all design measures $\xi \in \Xi$, where Ξ is the set of all probability measures on $I\!R$. Then

$$M(\xi, \underline{\theta}) = \int I(z, \underline{\theta}) d\xi(z),$$

where $I(z, \underline{\theta})$ is the Fisher information at a single point. In this context the solution to this continuous problem is in fact a design with discrete finite support though the real weights λ_i may not satisfy the integer restrictions. It is hoped that this optimal approximate design will be near the optimal discrete solution (see Silvey (1983)).

The choice of design criterion depends on the goal of the experiment. Sometimes the η-percentile of Ψ, μ_η, is of particular interest. In the parametrization of (1.1) and (1.2) $\mu_{50} = \mu$. If so, a natural design criterion called c-optimality is to minimize the asymptotic variance of the percentile estimate, $AV(\hat{\mu}_\eta)$, which turns out to be a one-dimensional linear function of the inverse of the information matrix. Note that 'minimizing' can be replaced by 'maximizing the negative'. An alternative criterion in this case suggested by Finney (1971) for μ (see also Abdelbasit and Plackett (1983)) is to minimize the squared half-length of a Fieller interval. If estimation of μ and β are of equal interest, various optimality criteria have been suggested. Some examples are A-, D- and E-optimality which entail minimizing by choice of design, the trace, the determinant and the maximum eigenvalue of the inverse of the information matrix, respectively. In this setting, the optimal design usually has 1-3 design points (see Wu (1988); Ford, Torsney and Wu (1992); Sitter and Wu (1993); Sitter and Fainaru (1997)).

There are a number of major concerns with this approach to designing experiments: (1) often "good" initial estimates of μ and β are not available and these 1-3 point designs are not robust to poor initial values, that is, if the initial values are incorrect, we may get very poor or even no estimates of μ and β; (2) 1-3 point designs may not allow adequate model checking; and (3) the choice of optimal design depends on the assumed model, which may be incorrect. For these reasons, usually the optimal design is used only as a benchmark to which to compare more heuristically chosen designs which suffer less from these difficulties. There have been some attempts to incorporate the initial lack of knowledge about the parameters into the design framework in a more automatic way (Chaloner and Larntz (1989); Sitter (1992)). These techniques, whether heuristic or more formal, generally produce designs which have more support points and are more spread out than the optimal designs, since they attempt to protect against incorrect initial values for μ and β. Though these designs protect against lack of knowledge about μ and β, they, in some sense, cannot make up for it. Sitter and Wu (1995) illustrate that often only a few of the design points lie in the central range of the response curve and thus give any information. Also, it may be that Ψ was incorrect. Thus the design may be very bad for this reason as well.

To address these difficulties, we must go back to the basic methodology of scientific enquiry. Postulate a hypothesis, collect observations, question or revise the hypothesis, collect more observations,... repeat until satisfied that the final hypothesis is correct. This procedure, or some variant of it, underlies almost any scientific investigation. In our context, this translates into the need to use more than one stage of experimentation. If the response can be observed in a short time, a fully sequential design is possible, where the next dose level, x_{n+1}, is chosen based on (x_i, r_i), $i = 1, \ldots, n$ (see Wu (1985); Young and Easterling (1994)). In many cases the time to response is too great to adopt such procedures. Sitter and Wu (1995), motivated by practical issues in drug development within the pharmaceutical industry, adopt the basic philosophy that one should use two-stages (or possibly three) of experimentation in many experiments of this type. Of course this philosophy may not apply in some cases. They give one possible general procedure for performing such a two-stage experiment and give some numerical results to illustrate the gains under various scenarios. Their basic strategy is to use a robust design with 3-5 dose levels for the first stage, using some portion of the affordable observations. Then analyze the data to get parameter estimates. Now treat these first stage estimates as the true values and choose a second stage design so that the combined design is "good". Note that if the first stage data indicates a different model than was first assumed is more appropriate, the second stage can be designed using this new model. In this paper, we assume that the first stage design has been obtained, run and analyzed, and the question under consideration is how to optimally choose a second stage design. Note that the first stage design itself could consist of a number of stages so the results apply to multi-stage designs as well.

The notion of two-stage or multi-stage designs is not new, especially in linear models (see Box, Hunter and Hunter (1978), Ch. 1 for discussion). For example, it is a fundamental aspect of response surface methodology (Box, Hunter and Hunter (1978), Ch. 15) and is often advocated as a means of breaking confounding in fractional factorials (Box, Hunter and Hunter (1978), p. 389-390). In non-linear models the situation is more complex since, unlike linear models, the information matrix involves the unknown parameters. For this reason less is known (see Minkin (1987)). However, it is this reason which makes the potential gains much greater, since a one-stage design must assume "good" initial values of these unknown parameters before data is collected.

The paper is organized as follows. Section 2 introduces the general problem and uses geometrical arguments to show that, for a class of Ψ which contains the logit and probit models and for a class of optimality criteria, the optimal second stage design will consist of two points symmetrically placed about the ED50 with possibly different weights at each point. In Section 3, we show that many of the design criteria in the literature belong to this class. In Section 4, as examples, we characterize the resulting D-, c-, E- and A-optimal second stage designs and reduce the problem further to a one variable maximization. We go on to give a small numerical example to illustrate the differences in the various optimality criteria. In Section 5, we discuss various extensions and make a few closing remarks.

2. Induced Design Space

Considering $M(\xi, \underline{\theta})$ with $\underline{\theta}$ fixed and known, we can view the optimality criterion as a function $\phi(\underline{t})$ depending on $\xi \in \Xi$ through $\underline{t} = (u, v, w)' = (S_0(\xi), S_1(\xi), S_2(\xi))'$, where $S_j(\xi) = \int z^j \psi(z) d\xi(z)$. This induces a design space $\mathcal{CH} = \{\underline{t} = (u, v, w)' = (S_0(\xi), S_1(\xi), S_2(\xi))' : \xi \in \Xi\}$. Consider the curve in $I\!R^3$,

$$\mathcal{H} = \{\underline{t} = (u, v, w)' : u = \psi(z), v = z\psi(z), w = z^2\psi(z), -\infty < z < \infty\},$$

and assume $z^2\psi(z)$ is bounded. This is true for many common models including the logit and probit (Wu (1988)). By looking at the form of $S_j(\xi)$, we see that $\mathcal{CH}$ is the convex hull of $\mathcal{H}$. Thus any design point $\xi \in \Xi$ implies a point $(u, v, w)' \in \mathcal{CH}$, while any point in $\mathcal{CH}$ can be generated by some design. Viewing $\mathcal{H}$ and $\mathcal{CH}$ is a very common geometric approach to optimal design for linear models (Silvey (1980)). Figure 1 shows the line $\mathcal{H}$ for the logit model. The line $\mathcal{H}$ lies on the upper and lower ridge of the figure and the shaded area represents what we term the *vertical boundary* of $\mathcal{CH}$. Such a vertical boundary exists for any Ψ such that Ψ' is symmetric about zero, since $-\{z\psi(z)\} = \{-z\}\{\psi(-z)\}$. Caratheodory's theorem (see Silvey (1980), p. 72) implies that any point in $\mathcal{CH}$ can be represented by a convex combination of a finite number of points in $\mathcal{H}$. Thus we need only consider designs with a finite number of support points.

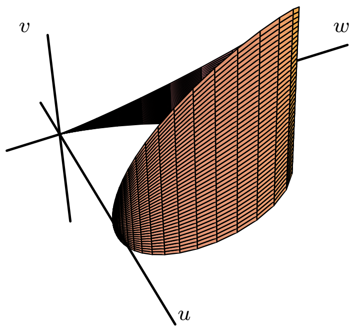

Figure 1. The curve $\mathcal{H}$ for the logit model.

Let

$$\xi_I = \{(\lambda_{Ii}, z_{Ii}) : \; i = 1, \ldots, k_I\}$$

be the first stage design, where $z_{Ii} = \hat{\beta}_I(x_{Ii} - \hat{\mu}_I)$ and x_{Ii} are the k_I first stage design points with weights $\lambda_{Ii} = n_{Ii}/n_I$ in obvious notation, and $(\hat{\beta}_I, \hat{\mu}_I)$ are the maximum likelihood estimates (mle) of the parameters from the first stage data. The problem now reduces to considering second stage designs of the form

$$\xi_{II} = \{(\lambda_{IIi}, z_{IIi}) : \; i = 1, \ldots, k_{II}\},$$

in similar notation, where $z_{IIi} = \hat{\beta}_I(x_{IIi} - \hat{\mu}_I)$ and n_{II} is considered fixed. The combined design matrix can then be written

$$M(\xi_{II}) = \begin{pmatrix} \hat{\beta}_I^2 S_0 & -S_1 \\ -S_1 & \frac{1}{\hat{\beta}_I^2} S_2 \end{pmatrix},$$

where $S_j = \epsilon S_{Ij} + (1-\epsilon) S_{IIj}$, $\epsilon = n_I/n$, $n = n_I + n_{II}$, $S_{Ij} = \sum_{i=1}^{k_I} \lambda_{Ii} z_{Ii}^j \psi(z_{Ii})$, and $S_{IIj} = \sum_{i=1}^{k_{II}} \lambda_{IIi} z_{IIi}^j \psi(z_{IIi})$ for $j = 0, 1, 2$.

In this framework, $\phi(\cdot)$ depends on x_{II} through $\underline{t} = \epsilon \underline{t}_I + (1-\epsilon)\underline{t}_{II}$ where $0 < \epsilon < 1$ is fixed, $\underline{t}_I \in \mathcal{CH}$ is fixed and $\underline{t}_{II} = (u, v, w)' = (S_{II0}(\xi_{II}), S_{II1}(\xi_{II}), S_{II2}(\xi_{II}))'$

$\in \mathcal{CH}$. So we wish to find $\underline{t}_{II} \in \mathcal{CH}$ which maximizes $\phi(\underline{t})$ and then find a design ξ_{II} which implies the resulting $\underline{t}$.

To do this, we need the following Lemma:

Lemma 1. *If $\underline{t}_I \in \mathcal{CH}$ is fixed and $\underline{t}_{II1}, \underline{t}_{II2} \in \mathcal{CH}$ are two other distinct points anywhere in $\mathcal{CH}$, then the line joining*

$$\underline{t}_1 = \epsilon \underline{t}_I + (1-\epsilon)\underline{t}_{II1} \in \mathcal{CH} \quad \text{and} \quad \underline{t}_2 = \epsilon \underline{t}_I + (1-\epsilon)\underline{t}_{II2} \in \mathcal{CH}$$

is parallel to the line joining $\underline{t}_{II1}$ and $\underline{t}_{II2}$, for any fixed real $0 < \epsilon < 1$.

Proof. First $\underline{t}_1, \underline{t}_2 \in \mathcal{CH}$ by convexity. The rest of the proof follows straightforwardly using similar triangles (see Forbes (1994)).

We can now state the main result:

Result 1. *Assume:*
(i) $\phi(\underline{t})$, $\underline{t} = (u, v, w)'$, *is non-decreasing in w for fixed u and v.*
(ii) *The curve $\{(u, v) :\ u = \psi(z),\ v = z\psi(z), z > 0\}$ in $I\!R^2$ is concave.*
(iii) *The curve $\{(u, w) :\ u = \psi(z),\ v = z^2\psi(z), z > 0\}$ in $I\!R^2$ is concave.*
For any fixed real $0 < \epsilon < 1$, if

$$\phi(\epsilon \underline{t}_I + (1-\epsilon)\underline{t}^*_{II}) = \max_{\underline{t}_{II} \in \mathcal{CH}} \phi(\epsilon \underline{t}_I + (1-\epsilon)\underline{t}_{II}),$$

then $\underline{t}^*_{II} \in \mathcal{VB}$, where $\mathcal{VB} = \{\underline{t} = \alpha \underline{t}_1 + (1-\alpha)\underline{t}_2 :\ \underline{t}_1 = (u, v, w)' \in \mathcal{H},\ \underline{t}_2 = (u, -v, w)' \in \mathcal{H}, 0 \le \alpha \le 1\}$ is the vertical boundary of $\mathcal{CH}$.

Proof. (i) together with Lemma 1 imply that given any $\underline{t}_{II} = (u_{II}, v_{II}, w_{II})' \in \mathcal{CH}$, we can increase $\phi(\epsilon \underline{t}_I + (1-\epsilon)\underline{t}_{II})$ by holding u_{II} and v_{II} fixed and increasing w_{II} until $\underline{t}_{II}$ hits the boundary of $\mathcal{CH}$. (ii) and (iii) insure that this resulting boundary point will always be in $\mathcal{VB}$.

Let us consider Result 1. First, we show in Section 3 that (i) holds for all of the main optimality criteria in the literature, in particular all of those considered in Silvey (1980). One can also show that (i) holds for F-optimality provided n is large. To consider (ii) and (iii), recall that we are restricting to symmetric $\psi(\cdot)$. Let the slope of the curve in (ii) be

$$\frac{\partial}{\partial z}\{z\psi(z)\} / \frac{\partial}{\partial z}\psi(z) = z + \psi(z)/\psi'(z) = z/2 + r(z)/2,$$

where $r(z) = z + 2\psi(z)/\psi'(z)$ appears in Wu (1988) equation (9) in his discussion of c-optimality. Assumption (ii) is implied by $r'(z) > 0$ for $z > 0$ and Wu proves that this holds for a number of $\psi(\cdot)$ including the logit and probit models. Let the slope of the curve in (iii) be

$$h(z) = \frac{\partial}{\partial z}\{z^2\psi(z)\} / \frac{\partial}{\partial z}\psi(z) = zr(z)$$

which appears in Sitter and Wu (1993), p. 333 in their discussion of F-, D-, and A-optimality. Assumption (iii) is implied by $h'(z) > 0$ for $z > 0$. Sitter and Wu (1993) show this holds for the logit model and note that though they have no rigorous proof, a plot of the curve suggests it also holds for the probit model.

It is easy to see that one could generalize Result 1 by removing assumptions (ii) and (iii) and stating as the conclusion that the optimal point lies on the boundary of the convex hull. This however would greatly reduce its usefulness. The power of Result 1 lies in the fact that by its very definition, any point in $\mathcal{VB}$ can be written as a convex combination of the two points on $\mathcal{H}$ which lie directly above and below it. Thus the search for the optimal second stage design can be restricted to designs of the form $\{(\lambda, z), (1-\lambda, -z) :\ 0 \leq \lambda \leq 1,\ z \geq 0\}$. This reduces the optimization problem greatly and in many cases allows some characterization of the designs.

3. Optimality Criteria

Silvey (1980) discusses a large list of optimality criteria which have been suggested in the literature, all of which share the following property. If we consider the design criterion as a function $\phi(M)$, where M is a non-negative definite matrix (i.e. the Fisher information matrix), then if M_1, M_2 and $M_1 - M_2$ are non-negative definite, then $\phi(M_1) \geq \phi(M_2)$. This list includes A-, D-, D_A-, D_s-, G-, E- and c-optimality.

This relates to assumption (i) of Result 1 in the following way. Take any two points in $\mathcal{CH}$ such that $\underline{t}_1 = (u, v, w_1)$ and $\underline{t}_2 = (u, v, w_2)$ with $w_1 \geq w_2$. Since they are in $\mathcal{CH}$ there exist two designs with Fisher information matrices

$$nM_i = n \begin{pmatrix} \beta^2 S_0 & -S_1 \\ -S_1 & \frac{1}{\beta^2} S_{2i} \end{pmatrix}$$

for $i = 1, 2$ such that $u = S_0$, $v = S_1$ and $w_i = S_{2i}$. We can then rewrite our optimality function, $\phi(\underline{t}_i)$ as a function $\phi^*(M_i)$, for $i = 1, 2$, since the β is a constant, so that $\phi(\underline{t}_i) = \phi^*(M_i)$. Now note that $w_1 \geq w_2$ implies $M_1 - M_2$ is non-negative definite. Thus for any of the optimality criteria in Silvey (1980), $\phi(\underline{t}_1) = \phi^*(M_1) \geq \phi^*(M_2) = \phi(\underline{t}_2)$ and thus (i) is satisfied.

Sitter and Fainaru (1997) discuss optimal design using the squared half-length of a Fieller interval for the EDη, i.e. F-optimality. They show that this criterion also satisfies (i), their Condition 3, and thus Result 1 applies.

4. Example Characterizations of Optimal Designs

4.1. D-optimal second stage design

Assuming that Ψ satisfies the conditions (ii) and (iii) of Result 1 and Ψ' is symmetric about zero, we need only consider two points, (λ, z) and $(1-\lambda, -z)$

for the second stage D-optimal design. Thus the problem has been reduced to maximizing

$$\{\epsilon S_{I0} + (1-\epsilon)\psi(z)\}\{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\} - \{\epsilon S_{I1} + (1-\epsilon)(2\lambda - 1)z\psi(z)\}^2$$

over $z > 0$ and $0 \le \lambda \le 1$, where $\epsilon = n_I/n$. The only term containing λ is the last one. If z is held fixed, the determinant will be maximized if the last term is minimized with respect to λ. That is, for a fixed $z > 0$, choose λ to minimize $\{\epsilon S_{I1} + (1-\epsilon)(2\lambda-1)z\psi(z)\}^2$. Taking the derivative with respect to λ and setting it to zero yields

$$\lambda = \lambda(z) = \frac{1}{2}\left(1 - \frac{\epsilon S_{I1}}{(1-\epsilon)z\psi(z)}\right).$$

Since $0 \le \lambda \le 1$ we have

$$\lambda = g(z) = \begin{cases} 0, & \text{if } \lambda(z) < 0, \\ \lambda(z), & \text{if } 0 \le \lambda(z) \le 1, \\ 1, & \text{if } \lambda(z) > 1. \end{cases} \tag{4.1}$$

Therefore our problem has now been reduced to a one variable non-linear optimization. We need to maximize

$$D(z) = \{\epsilon S_{I0} + (1-\epsilon)\psi(z)\}\{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\} - \{\epsilon S_{I1} + (1-\epsilon)(2g(z)-1)z\psi(z)\}^2$$

over $0 < z < \infty$. If z^* is the solution, the D-optimal second stage design is $\{(g(z^*), z^*), (1 - g(z^*), -z^*)\}$.

This characterization also gives us some insight. In the case where the first stage design was in fact nearly symmetric about $\hat{\mu}_I$, i.e. $S_{I1} \doteq 0$, the optimal second stage design will also be symmetric about $\hat{\mu}_I$, that is $\lambda \doteq 1/2$. Now consider the case where the first stage design has $S_{I1} < 0$ (i.e. the first stage design is skewed left) then $\lambda = g(z^*) > 1/2$ which implies a second stage design which is skewed to the right. Similarly, a first stage design with $S_{I1} > 0$ (i.e. skewed right) will have $\lambda = g(z^*) < 1/2$ which implies a second stage design which is skewed left. Thus, heuristically, if our original design covered only one side of the response curve, the second stage design should be weighted more toward the opposite side of the response curve.

4.2. c-optimal second stage design

For simplicity, we will consider the case where $\mu = \mu_{50}$ is of particular interest, and consider minimizing the $AV(\hat{\mu})$. This is a special case of c-optimality. Assuming that Ψ satisfies the conditions (ii) and (iii) of Result 1 and Ψ' is symmetric about zero, we need only consider two points, (λ, z) and $(1-\lambda, -z)$ for the second stage c-optimal design. In the one-stage c-optimal design problem we must worry about the possible singularity of the information matrix when a one-point design is used, since the c-optimal design in this special case is a one-point design. However in the two-stage context the first stage already has some design points and thus this difficulty is usually avoided when considering the combined design. Thus the problem has been reduced to minimizing

$$\frac{\{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\}}{[\{\epsilon S_{I0} + (1-\epsilon)\psi(z)\}\{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\} - \{\epsilon S_{I1} + (1-\epsilon)(2\lambda-1)z\psi(z)\}^2]}$$

over $z > 0$ and $0 \le \lambda \le 1$, where $\epsilon = n_I/n$. Using the results of the previous section, minimizing this with respect to $0 \le \lambda \le 1$ yields (4.1), since λ only appears in the last term of the denominator. Therefore our problem has now been reduced to a one variable non-linear optimization. We need to minimize

$$C(z) = \frac{\{\epsilon S_{I2} + (1-0\epsilon)z^2\psi(z)\}}{[\{\epsilon S_{I0} + (1-\epsilon)\psi(z)\}\{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\} - \{\epsilon S_{I1} + (1-\epsilon)(2g(z)-1)z\psi(z)\}^2]}$$

over $0 < z < \infty$. If z^* is the solution, the c-optimal second stage design is $\{(g(z^*), z^*), (1-g(z^*), -z^*)\}$. As with D-optimality, if our original design was skewed to one side of the response curve, the second stage design will be skewed toward the opposite side.

4.3. E-optimal second stage design

Using Result 1, we need only consider two points, (λ, z) and $(1-\lambda, -z)$ for the second stage E-optimal design; the design which maximizes the minimum eigenvalue of $M(\xi, \underline{\theta})$. Using $\det[M(\xi, \underline{\theta}) - \alpha I] = 0$, we obtain the eigenvalues

$$\alpha = (\hat{\beta}_I^2 S_0 + S_2/\hat{\beta}_I^2)/2 \pm \{(\hat{\beta}_I^2 S_0 - S_2/\hat{\beta}_I^2)^2 + 4S_1^2\}^{1/2}/2,$$

where $S_0 = \epsilon S_{I0} + (1-\epsilon)\psi(z)$, $S_1 = \epsilon S_{I1} + (1-\epsilon)(2\lambda-1)z\psi(z)$, and $S_2 = \epsilon S_{I2} + (1-\epsilon)z^2\psi(z)$. Since each term is positive, the minimum is

$$\alpha = (\hat{\beta}_I^2 S_0 + S_2/\hat{\beta}_I^2)/2 - \{(\hat{\beta}_I^2 S_0 - S_2/\hat{\beta}_I^2)^2 + 4S_1^2\}^{1/2}/2$$

and finding the E-optimal design reduces to maximizing

$$f(\lambda, z) = (1-\epsilon)\psi(z)(1 + z^2/\hat{\beta}_I^4) - \Big[\{\epsilon(S_{I0} - S_{I2}/\hat{\beta}_I^4) + (1-\epsilon)\psi(z)(1-z^2/\hat{\beta}_I^4)\}^2 + 4\{\epsilon S_{I1} + (1-\epsilon)(2\lambda-1)z\psi(z)\}^2/\hat{\beta}_I^4\Big]^{1/2}$$

over $0 \le \lambda \le 1$ and $z > 0$. Clearly, this is maximized over $0 \le \lambda \le 1$ for fixed z by choosing λ to minimize the last term under the square root. This again yields (4.1) and the problem reduces to finding $z > 0$ to maximize $E(z) = f(g(z), z)$, where $g(z)$ is defined in (4.1). If z^* is the solution, the optimal design will then be

$\{(g(z^*), z^*), (1-g(z^*), -z^*)\}$. So the the E-optimal second stage design problem is reduced to a one variable maximization.

4.4. A-optimal second stage design

Using Result 1, we need only consider two points (λ, z) and $(1-\lambda, -z)$ for the second stage A-optimal design the design which minimizes the trace of $M^{-1}(\eta, \underline{\theta})$. The problem reduces to minimizing

$$\frac{\hat{\beta}_I^2\{\epsilon S_{I0} + (1-\epsilon)\psi(z)\} + \{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\}/\hat{\beta}_I^2}{[\{\epsilon S_{I0} + (1-\epsilon)\psi(z)\}\{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\} - \{\epsilon S_{I1} + (1-\epsilon)(2\lambda-1)z\psi(z)\}^2]}$$

over $z > 0$ and $0 \leq \lambda \leq 1$. Using the same argument as in the previous three sections, minimizing this with respect to λ yields (4.1), since λ only appears in the last term of the denominator. Therefore our problem reduces to minimizing

$$A(z) = \frac{\hat{\beta}_I^2\{\epsilon S_{I0} + (1-\epsilon)\psi(z)\} + \{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\}/\hat{\beta}_I^2}{[\{\epsilon S_{I0} + (1-\epsilon)\psi(z)\}\{\epsilon S_{I2} + (1-\epsilon)z^2\psi(z)\} - \{\epsilon S_{I1} + (1-\epsilon)(2g(z)-1)z\psi(z)\}^2]}$$

over $z > 0$, where $g(z)$ is defined in (4.1). If z^* is the solution, the optimal design will then be $\{(g(z^*), z^*), (1-g(z^*), -z^*)\}$. So the the A-optimal second stage design problem is reduced to a one variable maximization.

Similar characterizations can be done for various optimality criteria, though the algebra can be quite tedious. For some guidance, one can look at the characterizations of F-optimal designs in Sitter and Wu (1993) and of F- and G-optimal designs in Sitter and Fainaru (1997).

4.5. A numerical illustration

To illustrate the presented ideas we consider two hypothetical scenarios. Let us pretend that an experimenter wishes to consider a two stage experiment in the stated context. From previous experience, they believe the logit model is applicable with dose measured on the log scale. It is also believed that μ and β will be near 5 and 0.5, respectively. At least, if forced to pick only point estimates, these are the values the experimenter would give. In scenario 1, we assume the resources available allow $n = 60$ subjects to be used, while in scenario 2 $n = 100$ subjects are available. In scenario 1, suppose the experimenter wishes to do the experiment in two stages using $n_I = 20$ subjects in the first stage and $n_{II} = 40$ in the second, while in scenario 2, $n_I = 20$ and $n_{II} = 80$. We also suppose that after some consideration of how confident they were in the model and their parameter values, and the need for model checking, they decided to place $n_i = 5$ subjects at each of

$$x_1 = 2, \quad x_2 = 4, \quad x_3 = 6, \quad x_4 = 8,$$

where x =log dose, in both scenarios. For this illustration, it is not important how this design was reached, however, in fact, a more detailed scenario which results in this design using the method of robust design in Sitter (1992) is given in Forbes (1994), Ch. 3 and Sitter and Wu (1995).

Now suppose the true parameter values were actually $\mu = 3$ and $\beta = .5$. That is, the experimenter was incorrect in location, thus the resulting design is skewed right. To illustrate how the various optimality criteria would pick a second stage, we suppose that the resulting response vector was the same in each scenario; $r = (2, 3, 4, 5)'$, which implies the mle's $\hat{\mu}_I = 3.015$ and $\hat{\beta}_I = 0.59$, and the logit model fits the data well. Note that it does not matter if the first stage design was obtained assuming a logit model or not provided the data is well fit by a logit model. We can now use the theory of the previous section to obtain the D-, $c-$, E- and A-optimal second stage designs for each scenario.

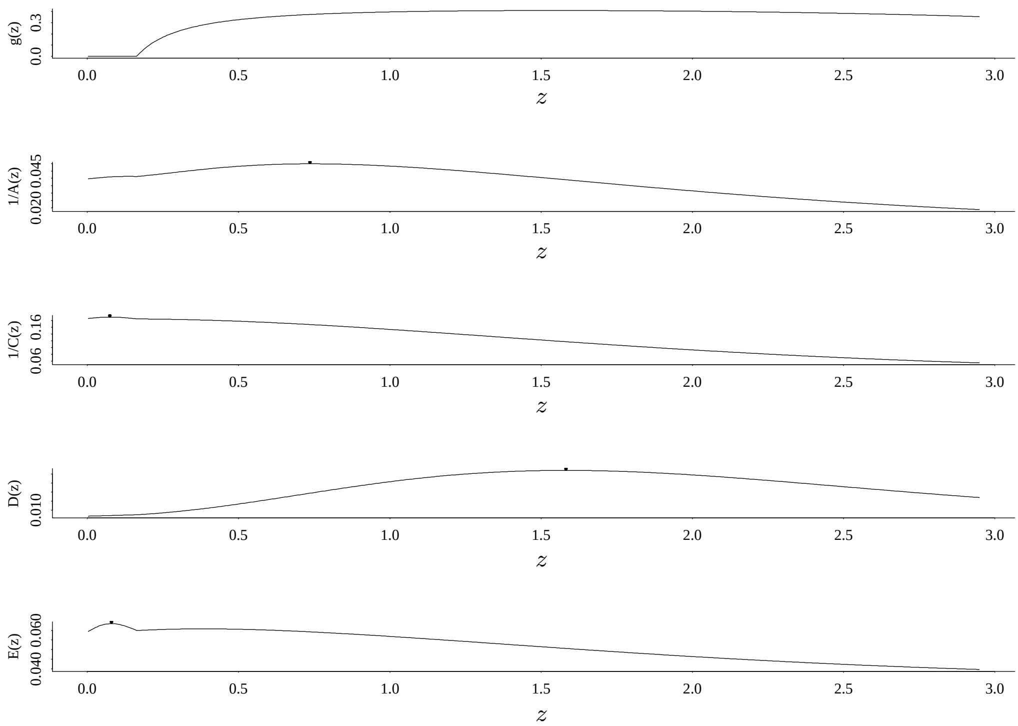

Figure 2. Plots of $g(z)$, $1/A(z)$, $1/C(z)$, $D(z)$ and $E(z)$ for scenario 1.

First consider scenario 1, where $n_I = 20$ and $n_{II} = 40$, which implies $\epsilon = 1/3$. In Figure 2, we plot $g(z)$, $1/A(z)$, $1/C(z)$, $D(z)$ and $E(z)$ for $0 \leq z = \hat{\beta}_I(x - \hat{\mu}_I) <$

3.0. The maximum, z^*, for each of the optimality criteria is clearly marked, and if one moves vertically up from the maximum to the plot of $g(z)$ one can ascertain the value of $\lambda^* = g(z^*)$ for the optimal design. We see immediately that the A- and D-optimal second stage designs have two points, while the $c-$ and E-optimal second stage designs have only one. Specifically, the $(1-\lambda^*, -z^*)$, (λ^*, z^*) are:
A-optimal, $(0.63, -0.74)$, $(0.37, 0.74)$;
D-optimal, $(0.59, -1.58)$, $(0.41, 1.58)$;
c-optimal, $(1, -0.07)$;
E-optimal, $(1, -0.08)$.
Figure 3, depicts the resulting combined first and second stage design for each criterion. We see that in all cases the second stage design is more heavily weighted to the left of $\hat{\mu}_I = 3.015$ to correct the asymmetry in the first stage design.

Viewing Figure 2 once more we can see that $D(z)$ appears to have only one local maximum and is well behaved, but $1/A(z)$, $1/C(z)$ and $E(z)$ show the possibility of more than one maximum. Thus one should be careful when routinely applying a maximization routine. It is much simpler to just plot the function.

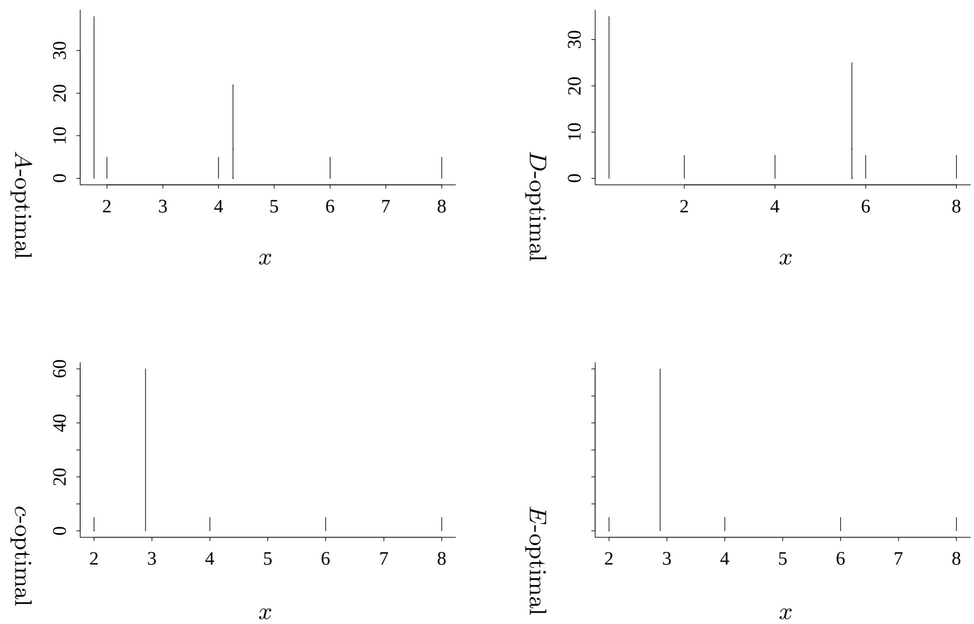

Figure 3. The resulting A-, c-, D- and E-optimal combined first and second stage designs for scenario 1.

We introduce scenario 2 to illuminate a rather surprising result in scenario 1. In scenario 1, it is not surprising that the c-optimal second stage design has only one point, while the A- and D-optimal second stage designs have two points, since: (i) the A- and D-criterion were derived to measure the accuracy of both the estimate of μ and β, while the c-criterion is only concerned with μ; and (ii) the A- and D-optimal one stage designs for this model have two points (see Sitter and Wu (1993)) while the c-optimal one stage design has only one point (see Wu (1988)). The surprising result is that the E-optimal second stage design has only one point and closely resembles the c-optimal second stage design instead of the A- and D-optimal second stage designs. The E-criterion was derived to measure the accuracy of both parameters and the E-optimal one stage design for this model has two points (see Sitter and Fainaru (1997)).

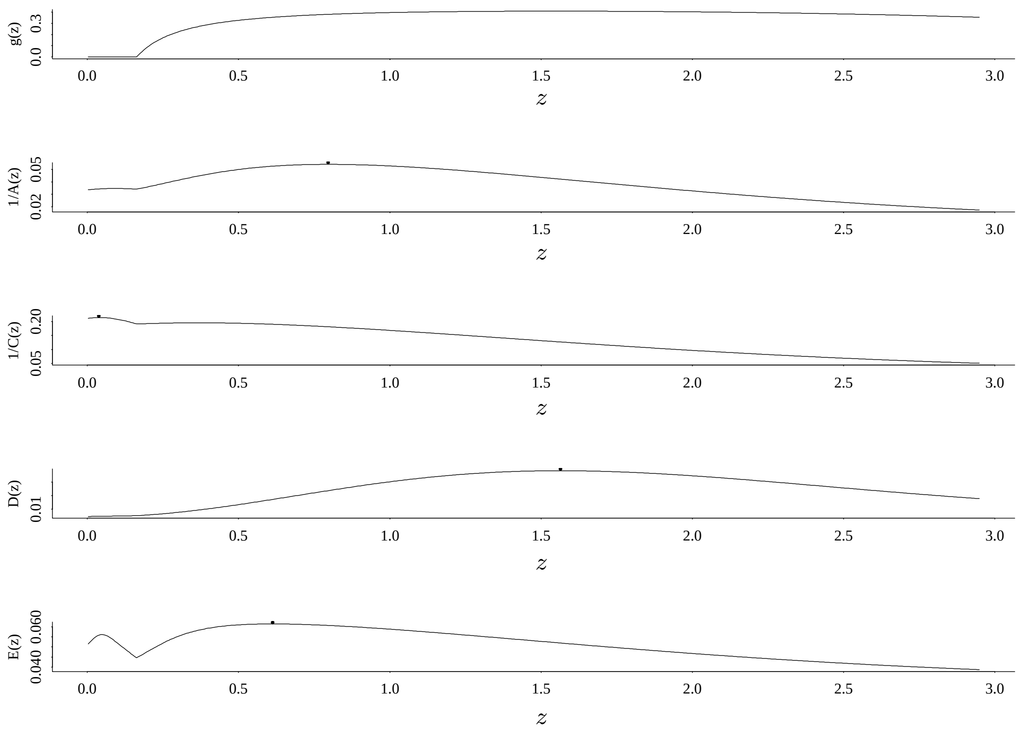

Figure 4. Plots of $g(z)$, $1/A(z)$, $1/C(z)$, $D(z)$ and $E(z)$ for scenario 2.

In scenario 2, $n_I = 20$ and $n_{II} = 80$, so that $\epsilon = 1/5$. In Figure 4 we again plot $g(z)$, $1/A(z)$, $1/C(z)$, $D(z)$ and $E(z)$. From this we see that the curves $1/C(z)$, $1/A(z)$ and $D(z)$ do not change much from scenario 1, but the $E(z)$ curve is fundamentally different. In Figure 2, $E(z)$ had two local maxima with the left one dominant, but in Figure 4 the right one has become dominant.

This implies a switch from a one to a two point E-optimal second stage design. Specifically, the $(1-\lambda^*, -z^*)$, (λ^*, z^*) are:
A-optimal, $(0.62, -0.80)$, $(0.38, 0.80)$;
D-optimal, $(0.59, -1.56)$, $(0.41, 1.56)$;
c-optimal, $(1, -0.04)$;
E-optimal, $(0.65, -0.61)$, $(0.35, 0.61)$
and the E-optimal second stage design more closely resembles the A- and D-optimal second stage designs.

Viewing Figure 2 and 4 we might anticipate that $1/C(z)$ and $1/A(z)$ might also display this switching property from one to two point designs, since they display a potential for two local maxima. We have investigated various scenarios and have found that the A-criterion does sometimes yield a one or two point design, while the c-criterion seems to always yield only one point. We should also note that in all cases considered the second stage design is skewed left, symmetric and skewed right precisely opposite to the original designs and thus creates a more symmetric combined design.

5. Conclusion

In this article, we have reduced the problem of obtaining an optimal second stage design for a class of binary response models to a two variable (in many useful cases a one variable) maximization problem for almost any optimality criteria. This class of models includes the two most common models: the logit and the probit. There are of course many questions one can ask and extensions one would like to have. We make a few comments in this regard.

Often the experimenter would like to restrict the dose levels within a bounded range. This amounts to removing a section of $\mathcal{VB}$ (see Figure 1). If the bound on x is symmetric about the $\hat{\mu}_I$, that is, the bound amounts to $z \in [-d, d]$ for some $d > 0$, then the problem is no more difficult and a similar result to Result 1 is immediately available. However, if this is not so which will usually be the case, characterization of $\mathcal{CH}$ is more difficult. There has been some work done on this for one stage designs (see Ford, Torsney and Wu (1992)). It is possible that such extensions could be done here, though they would be more difficult.

Conditions (i) and (ii) which we place on $\psi(\cdot)$ restrict our model to case (i) of Sitter and Wu (1993). In their case (ii) they relax Condition (ii). If we do this, Result 1 still holds except that $\mathcal{VB}$ will no longer be vertical. It will consist of a vertical curved planar piece with a sloped roof on top and a sloped floor at the bottom. This means that Result 1 would have reduced usefulness since we could not easily characterize designs which generate points on $\mathcal{VB}$. In the one stage design problem it is sometimes possible in specific cases to prove that the resulting point will in fact lie on the vertical part of $\mathcal{VB}$ and thus the resulting design would consist of 3 points symmetrically placed about the ED50 with possibly unequal weighting (see Sitter and Wu (1993)), however in a two stage design a bad first stage will effect the resulting second stage point and even in these special cases may force it outside of the vertical part of $\mathcal{VB}$.

Acknowledgement

The first author's research was supported by a grant from the Natural Sciences and Engineering Council of Canada. We thank the referees for their suggestions.

References

Abdelbasit, K. M. and Plackett, R. L. (1983). Experimental design for binary data. *J. Amer. Statist. Assoc.* **78**, 90-98.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building.* John Wiley, New York.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inference* **21**, 191-208.

Finney, D. J. (1971). *Probit Analysis*, 3rd edition. Cambridge University Press, Cambridge.

Forbes, B. E. (1994). *Two-Stage Designs for Binary Response Data.* M.Sc. Thesis, Carleton University, Ottawa, Canada.

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.

Minkin, S. (1987). Optimal designs for binary data. *J. Amer. Statist. Assoc.* **82**, 1098-1103.

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

Sitter, R. R. (1992). Robust designs for binary data. *Biometrics* **48**, 1145-1155.

Sitter, R. R. and Fainaru, I. (1997). Optimal designs for the logit and probit models for binary data. *Canad. J. Statist.* **25**, 175-189.

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.

Sitter, R. R. and Wu, C. F. J. (1995). Two-stage design of quantal response studies. University of Michigan Technical Report.

Wu, C. F. J. (1985). Efficient sequential designs with binary data. *J. Amer. Statist. Assoc.* **80**, 974-984.

Wu, C. F. J. (1988). Optimal design for percentile estimation of a quantal response curve. In *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).

Young, L. J. and Easterling, R. G. (1994). Estimation of extreme quantiles based on sensitivity tests: a comparative study. *Technometrics* **36**, 48-60.

Department of Mathematics and Statistics, Simon Fraser University, Burnaby, B.C. V5A 1S6 Canada.

E-mail: sitter@math.sfu.ca

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

E-mail: barbara.e.forbes@ac.ca

(Received April 1995; accepted September 1996)