Science and Technology, 3 (1998) 57-66

Constant Information Designs for Binary Response Data

K. M. Abdelbasit

Department of Mathematics and Statistics, College of Science, Sultan Qaboos University, P.O.Box 36, Al-Khod 123, Muscat, Sultanate of Oman.

تصاميم المعلومات الثابتة لبيانات الاستجابة الثنائية

خضر محمد عبد الباسط

خلاصة : عند استخدام النماذج غير الخطية ، يعتمد تصميم التجربة على معالم النموذج غير المعروفة . ندرس في هذا البحث تصميم التجارب في حالة الاستجابة الثنائية ونعمم التصميمات ذات المعلومات الثابتة التي اقترحها فيشر (١٩٢٢) . ينتج من هذه التصميمات تباين ثابت . نثبت وجود هذه التصميمات في حالة النماذج ذات المعلم الواحد وندرس بناء هذه التصميمات ونختبر أدائها . عندما يعتمد النموذج على معلمين نثبت أنه يمكن تصميم التجربة بحيث يكون محدد مصفوفة المعلومات لا يعتمد على معالم النموذج . ندرس في هذه الحالة أيضا بناء التصميمات ونختبر أدائها

ABSTRACT: A major problem in designing experiments when the assumed model is nonlinear, is the dependence of the designs on the values of the unknown parameters. We consider in this article designs for binary data and generalize the constant information criterion suggested by Fisher (1922). The criterion calls for designs that achieve a specific proportion of the total constant information. This leads to designs where dependence of Fisher's information on the unknown parameters is very little, thus leading to constant variances. We show that such designs exist for any single parameter model, extending Fisher's result for the exponential model. We discuss the construction of such designs and investigate their performance as measured by the achievement of constant information. When two parameters are needed to specify the model, we show that experiments can be designed so that the determinant of the information matrix is independent of the parameters. Construction of designs and examining their performance are also investigated for the two parameter case.

Key Words. Constant information; Binary data; D-optimality; models for binary data;

Optimal design criteria for linear models based on Fisher's information matrix are used for non-linear models as well. A problem introduced by non-linearity of the model is that the information matrix, and hence the optimal design, depends on the unknown parameters. Non-linear design problems are reviewed by Cochran (1973) and by Ford, et al.(1989). Abdelbasit and Plackett (1981) reviewed the design problems when the response variable is categorical. In this paper we deal with the case where the response to a stimulus is binary (i.e response/no response). Abdelbasit and Plackett (1982) considered this problem when the stimulus is a mixture of two stimuli. For a single stimulus, and assuming a logistic model, Abdelbasit and Plackett (1983) derived D-optimal designs under the restriction of symmetry. Their work was generalized by Minkin (1987) and Khan and Yazidi(1988).

A number of methods are used to overcome the problem of dependence of designs on the unknown parameters. Among these are:

1. Initial estimates known: Here we use previous knowledge or guess to specify initial estimates for the parameters. Abdelbasit and Plackett(1983) investigated the robustness of such designs to poor initial estimates. Sitter (1992) obtained designs that are robust to poor initial estimates. Cox(1988) studied the case of estimating a small treatment difference and showed that the problem caused by nonlinearity is not severe and that the usual normal theory applies well if the data is not very heterogeneous.

2. Sequential designs: We start with initial estimates and proceed sequentially in a multistage design, with each stage providing initial estimates for the next. Abdelbasit and Plackett (1983) obtained analytical results for two stage designs using an exponential model. They also investigated the efficiency of multistage designs. Wu(1985) obtained a class of sequential designs for estimating the percentiles of the quantal response curve.

3. Bayesian Methods: Assume a prior distribution for

the unknown parameters and maximize the posterior expected information. Chaloner and Larntz (1989) derived optimal Bayesian designs for the logistic model.

4. Constant Information: Obtain designs that make the information function, or the determinant of the information matrix, constant. This is the subject matter of this paper. It continues the work of Abdelbasit and Plackett (1983).

Fisher (1922, 1966) considered the problem of estimating the density of organisms by the dilution method. Samples of liquid are taken and we observe whether they contain organisms (fertile) or not (sterile). At each dilution n plates are examined and the probability of a fertile plate is

$$p(x) = 1 - \exp(-\beta x)\ ;\ x > 0,\ \beta > 0 \tag{1.1}$$

where β is the unknown parameter and x corresponds to the dilution. The series of dilutions used form a geometric progression

$$x_{i+1} = ca^{-i}\ ;\ c > 0\ , a > 0, i = 0, 1, 2, \dots \tag{1.2}$$

where a and c are constants, with a usually 2 (Cochran 1973). The information about log β at dilution x is

$$I(\log\beta) = \frac{n(\beta x)^2}{\exp(\beta x) - 1} \tag{1.3}$$

The total information from n plates at each dilution is (see Plackett 1981, pp 59-61).

$$I(\log\beta) = \sum_{i=0}^{\infty} n(\beta ca^{-i})^2 / (\exp(\beta ca^{-i}) - 1) \tag{1.4}$$

Note that

$$I(\log\beta) = \beta^2 I(\beta).$$

Thus maximizing $I(\log\beta)$ is equivalent to minimizing the asymptotic coefficient of variation of the maximum likelihood estimator of β. Fisher (1966) argued (see also Cochran 1973) that it is the magnitude of the variance relative to the parameter, which should be minimized, rather than the variance in isolation. He proposed the minimization of the coefficient of variation, which explains taking $I(\log\beta)$ in (1.4) rather than $I(\beta)$. $I(\log\beta)$ in (1.3), is maximized at $\beta x = 1.59$. This has no practical value, since the choice of x that maximizes $I(\log\beta)$ depends on the valaue of unknown parameter β itself. Fisher (1922) noticed that the total information $I(\log\beta)$, given by (1.4), is almost independent of β and has an approximate value of ($\pi^2/6 \log a$). Thus experiments can be designed in such a way that the information function is independent of the parameter β. Such a property is desirable because the aymptotic dispersion matrix of the estimators remains the same (Abdelbasit and Plackett 1983). In what follows, we generalize Fisher's results for any single parameter (sections 3 and 4) and examine possible generalizations to the two parameter models (section 5).

General Formulation.

The general problem may be formulated as follows (Abdelbasit and Plackett 1983). The experimenter is investigating the relationship between the level of a stimulus and the rate of response to this stimulus. Of particular interest in many cases is the stimulus level that causes a prespecified response .eg. 50% or 90%. The total number of subjects N available for testing are divided into k groups. The n_i subjects in group i are given stimulus x_i (usually on a tranformed scale) and the number that respond R_i is observed. It is assumed that each subject has a threshold level of stimulus above which it responds with probability one. The random variable of the population thresholds is called the tolerance and has cumulative distribution function $\tilde{F}(x_i, \theta)$. Thus $R_1, R_2, \dots, R_k$ are mutually independent and R_i has a binomial distribution with index n_i and probability

$$p(x_i) = \tilde{F}(x_i, \theta)\ ;\ i = 1, 2, \dots, k \tag{2.1}$$

where θ is a vector of unknown parameters and $\tilde{F}(., \theta)$ a specified distribution function.

The design problem is to choose k, $\{x_i\}$ and $\{n_i\}$ in some optimal way, subject to the condition $\sum n_i = N$ is fixed. Optimality criteria are based on some measure of the information about θ provided by the experiment. For a single parameter, the usual criterion is to maximize Fisher's information function. This is equivalent to the minimization of the asymptotic variance of the maximum likelihood estimator of the parameter. When more than one parameter is needed to specify the model $\tilde{F}(x_i, \theta)$, optimality criteria call for the minimization of some real valued functions of the dispersion matrix. A number of optimality criteria were proposed (see Silvey 1980; chapter 2). The most intensively studied is D-optimality which calls for maximizing the determinant of the information matrix.

We deal here with the criterion of constant information, suggested in section 1. In what follows we use $f_d(.)$ to denote the function $f(.)$ evaluated on a discrete set of design points. We drop the subscript d when a continuous version of $f(.)$ is considered. Consider the model (2.1), where θ is a single parameter. Then

$$I_d(\theta)=\sum_{i=1}^{k}\frac{n_i(\partial\tilde{F}(x_i,\theta)/\partial\theta)^2}{\tilde{F}(x_i,\theta)\{1-\tilde{F}(x_i,\theta)\}} \quad (2.2)$$

where n_i subjects are tested at x_i ($i = 1, 2, \ldots, k$). For any function $\lambda(\theta)$ of θ, the information about $\lambda(\theta)$ is given by

$$I_d(\lambda(\theta))=\frac{I_d(\theta)}{(\lambda'(\theta))^2} \quad (2.3)$$

where

$$\lambda'(\theta)=\frac{d\lambda(\theta)}{d\theta}. \quad (2.4)$$

Our objective is to find $\{x_i\}$ and $\{n_i\}$ such that $I_d(\lambda(\theta))$ is as independent of θ as possible for some choice of $\lambda(.)$. Now (2.3) can be written as

$$I_d(\lambda(\theta))=N\sum_{i=1}^{k}\frac{n_i}{N}G(x_i,\theta)\{\lambda'(\theta)\}^{-2} \quad (2.5)$$

where

$$G(x,\theta)=\frac{(\partial\tilde{F}(x,\theta)/\partial\theta)^2}{\tilde{F}(x,\theta)\{1-\tilde{F}(x,\theta)\}} \quad (2.6)$$

Expression (1.4) for the dilution series follows on taking $\tilde{F}(x,\beta) = 1 - \exp(-\beta x)$, $n_i = n$, $\lambda(\beta) = \log\beta$ and $k \to \infty$.

Think of a random variable X with probability function $m_d(x)$ where

$$m_d(x)=P(X=x)=\begin{cases}\dfrac{n_i}{N} & x=x_i,\ i=1,2,\ldots,k\\ 0 & elsewhere\end{cases} \quad (2.7)$$

then

$$I_d(\lambda(\theta))=\frac{NE\{G(X,\theta)\}}{\{\lambda'(\theta)\}^2} \quad (2.8)$$

where E denotes expectation.

The problem is to find k, $\{x_i : i = 1, 2, \ldots, k\}$ and the probability function $m_d(.)$ which make (2.8) independent of θ for some choice of $\lambda(.)$. The discrete problem (2.8) does not have a general solution. The problem is simplified a bit if we replace the discrete distribution $m_d(x)$ by a continuous distribution with density $m(x)$. In this case (2.8) becomes

$$I(\lambda(\theta))=\frac{N}{(\lambda'(\theta))^2}\int_{-\infty}^{\infty}G(x,\theta)m(x)dx \quad (2.9)$$

Choices of $m(.)$ and $\lambda(.)$ that make (2.9) independent of θ are possible (Plackett 1981, pp 59-61). In the following sections we consider such designs and methods of contruction of finite approximations to the resulting continuous design measures.

A Single Scale Parameter

Suppose that

$$p(x)=\tilde{F}(\beta,x)=F(\beta x),\ \ \beta>0. \quad (3.1)$$

Then (2.5) gives

$$I_d(\lambda(\beta))=N\sum_{i=1}^{k}\frac{n_i}{N}x_i^2\,\omega(\beta x_i)/(\lambda'(\beta))^2 \quad (3.2)$$

where

$$\omega(t)=(f(t))^2/F(t)\{1-F(t)\} \quad (3.3)$$

and $f(t)=\frac{dF(t)}{dt}$ is the pdf corresponding to the cdf F(t). Since $p(x)$ is known when $x = 0$, only positive values of x are necessary. Thus (2.9) becomes

$$I(\lambda(\beta))=\frac{N}{(\lambda'(\beta))^2}\int_{0}^{\infty}x^2\omega(\beta x)m(x)dx. \quad (3.4)$$

Take

$$m(x)=A/x \quad (3.5)$$

where A is a constant, $\lambda(\beta) = \log\beta$ and substitute $t = \beta x$ to get

$$I(\log\beta)=NA\int_{0}^{\infty}t\omega(t)dt \quad (3.6)$$

which is independent of β. Fisher's result on the dilution method follows on taking $F(x) = 1 - \exp(-x)$.

The function $m(x)$ in (3.5) is not a proper distribution. In practice it can be used in one of two ways:

(1) When seeking design points $\{x_i\}$ we can approximate $m(x)$ by samples of equal size at points $\{x_i\}$ such that

$$\int_{x_i}^{x_{i+1}} m(x)dx = constant; i=1,2,...,k-1 \qquad (3.7)$$

To determine the constant we need to specify x_1, k and x_k. Specification of the first and last design points (x_1 and x_k) can be based on the available information on the parameter β as indicated below. Note, also, that the discrete approximation $\{x_i\}$ for $m(x)$ and $Km(x)$ is the same for any arbitrary constant K. Thus (3.5) is *approximated* by the same k-points design irrespective of the value of A.

(2) When the points $\{x_i : i = 1, 2, \ldots, k\}$ are fixed, sample sizes are determined by

$$\alpha_i = \frac{m(x_i)}{\sum_{j=1}^{k} m(x_j)}; \ i=1,2,..., k$$

where a proportion α_i of the subjects is tested at x_i.

Some knowledge of β may be available before the experiment. Suppose it is known that

$$\beta_L \leq \beta \leq \beta_H \qquad (3.8)$$

where β_L and β_H are given values. For any $\delta \in (0, 1)$ define $t(\delta)$ by (Plackett 1981)

$$\int_0^{t(\delta)} t\omega(t)dt = \delta \int_0^{\infty} t\omega(t)dt \qquad (3.9)$$

Let x_L and x_H be such that

$$\beta_L x_H = t(1-\epsilon) \ \text{ and } \ \beta_H x_L = t(\epsilon) \qquad (3.10)$$

where ϵ is small, and consider a design

$$m(x) = \begin{cases} A/x & x_L \leq x \leq x_H \\ 0 & elsewhere \end{cases} \qquad (3.11)$$

This ensures that

$$NA(1-2\epsilon) \int_0^{\infty} t\omega(t)dt \leq I(\log\beta) \leq NA \int_0^{\infty} t\omega(t)dt \qquad (3.12)$$

for all $\beta \in [\beta_L, \beta_H]$. Thus at least a proportion $(1 - 2\epsilon)$ of the constant total information is achieved. A k-point approximation to (3.11) has samples of equal sizes at points $\{x_i\}$ determined by (3.7), i.e., $\log(x_{i+1}/x_i)$ = constant. The points form a geometric progression. Setting $x_1 = x_L$, $x_k = x_H$, the points are

$$x_i = x_1 c^{i-1}; \ i=1,2,...,k \qquad (3.13)$$

where $c = \frac{1}{k-1} \log\left(\frac{x_H}{x_L}\right)$.

Recall that the discrete approximation of (3.11) given by (3.13) is the same irrespective of the value of A. A sensible choice of A is the one that makes $m(x)$ a density function.

Questions of interest are:

Q1. To what extent is $I_d(\log\beta)$, from the k - point design (eqn. 3.2) independent of β?

Q2. How many points are needed, for a k - point design to achieve this proportion of $(1 - 2\epsilon)$, of the total information?

These questions are addressed in the following example.

EXAMPLE 1: Take $-1 \leq \log\beta \leq 1$, $F(.)$ the logistic distribution function and $\epsilon = 0.05$. From (3.9), $t(0.05) \approx 0.6$, $t(0.95) \approx 5.2$ and $x_L = 0.221$, $x_H = 14.135$.

With samples of equal size at x_i and $\lambda(\beta) = \log\beta$, (3.2) becomes

$$I_d(\log\beta) = N(\beta^2 \sum_{i=1}^{k} \omega_i x_i^2/k) \qquad (3.14)$$

where

$$\omega_i = \omega(x_i) = \frac{\exp(\beta x_i)}{1+\exp(\beta x_i)}.$$

For k-point designs, obtained by (3.13), values of $I_d(\log\beta)/N$, given by (3.14) are calculated for $\log\beta = -1(0.2)1$ and the mean, standard deviation and the coefficient of

TABLE 1

$I(\log \beta)/N$ over $\log \beta = -1(0.2)\,1$.

K	Mean	S.D	Coefficient of Variation
2	0.0231	0.02120	0.917
3	0.1096	0.0313	0.286
4	0.1237	0.0078	0.063
5	0.1326	0.0016	0.012
6	0.1379	0.0011	0.008
7	0.1417	0.0013	0.009
8	0.1444	0.0015	0.010
9	0.1465	0.0016	0.011
10	0.1482	0.0017	0.012
11	0.1496	0.0018	0.012
12	0.1507	0.0019	0.013
13	0.1517	0.0020	0.013
20	0.1556	0.0024	0.015
30	0.1580	0.0026	0.017
50	0.1559	0.0029	0.018
100	0.1613	0.0031	0.019

variation of $I_d(\log \beta)/N$ are computed for different values of k. The results are presented in Table 1. From the table we see that

(i) $I_d(\log \beta)/N$ increases with k, but rather slowly from six points onward.

(ii) Minimum variability is achieved at $k = 6$. The coefficient of variation increases slightly with k for $k > 6$ but remains small.

From (3.6), we see that the total information about $\log\beta$ for the logistic distribution function is $NA \log 2$. Choice of $A = \left(\log \frac{x_H}{x_L}\right)^{-1}$ that makes (3.11) a density gives

$$I(\log\beta)/N = \log 2/\log\left(\frac{14.135}{0.221}\right) = 0.1667 \qquad (3.15)$$

with $\epsilon = 0.05$, $1 - 2\epsilon = 0.9$ and

$$(1-2\epsilon)\,\frac{I(\log\beta)}{N} = 0.1500 \qquad (3.16)$$

When the interval $[x_L, x_H]$ is reasonably covered, we expect the mean of $I_d(\log \beta)$ to exceed 0.1500. Table 1 shows that more than 10 points are needed to cover this range suitably.

A single location parameter

Similar results as those of section 3 hold for a location parameter where

$$p(x) = F(x-\mu) \qquad (4.1)$$

$$I_d(\lambda(\mu)) = \sum_{i=1}^{k} n_i \omega(x_i - \mu)/(\lambda'(\mu))^2 \qquad (4.2)$$

where $\omega(t)$ is defined by (3.3). The analogue of (3.4) is

$$I((\lambda(\mu)) = \frac{N}{(\lambda'(\mu))^2}\int_{-\infty}^{\infty} \omega(x-\mu)m(x)dx \qquad (4.3)$$

Take $m(x) = A$, $\lambda(\mu) = \mu$, then

$$I(\mu) = NA\int_{-\infty}^{\infty} \omega(t)dt \qquad (4.4)$$

indepedent of μ.

It follows from (3.6) and (4.4) that Fisher's result generalizes to any family with a single location or scale parameter. It is interesting to note the difference between the two cases. In the case of location parameter μ, designs which make the asymptotic variance of maximum likelihood estimator(*MLE*) $\hat{\mu}$, independent of μ exist. For a scale parameter β, designs which make the asymptotic coefficient of variation of the *MLE* $\hat{\beta}$, independent of β exist.

If it is known that

$$\mu_L \le \mu \le \mu_H, \qquad (4.5)$$

we can define $t(\delta)$ for $\delta \in (0, 1)$ by

$$\int_{-\infty}^{t(\delta)} \omega(t)dt = \delta\int_{-\infty}^{\infty} w(t)dt \qquad (4.6)$$

and let $x_L = \mu_L + t(\epsilon)$, $x_H = \mu_H + t(1-\epsilon)$ where ϵ is small. Then take

$$m(x) = \begin{cases} A & x_L \le x \le x_H \\ 0 & elsewhere \end{cases} \qquad (4.7)$$

giving

$$NA(1-2\epsilon)\int_{-\infty}^{\infty}\omega(t)dt \le I(\mu) \le NA\int_{-\infty}^{\infty}\omega(t)dt \qquad (4.8)$$

for $\mu\in[\mu_L, \mu_H]$. A k-point approximation to (4.7) has samples of equal sizes at points equally spaced between $x_1 = x_L$ and $x_k = x_H$. The value of A in (4.8) can be taken as $(x_H - x_L)^{-1}$. The stability of $I_d(\mu)$ and the proportion of the total information achieved from k-point designs are explored in the following example.

EXAMPLE 2: Take $-1 \le \mu \le 1$, $F(.)$ the logistic distribution function and $\epsilon = 0.05$. Then (4.6) gives $-t(0.05) = t(0.95) = 2.94 \sim 3$. With samples of equal sizes at x_i, (4.2) gives

$$I_d(\mu) = N\left\{\sum_{i=1}^{k}\omega(x_i-\mu)/k\right\} \qquad (4.9)$$

$I_d(\mu)/N$ is computed for $\mu = -1(0.2)1$, and different k-points designs equally spaced between $x = -4$ and 4. The mean, standard deviation and the coefficient of variation for different values of k are tabulated in Table 2. The table shows that

1. Minimum variability at $k = 6$ and the coefficient of variation stays small for $k > 6$.

2. $I_d(\mu)$ shows a lot less variability than $I_d(\log\beta)$.

Also from (4.4) with $A = 8$, we get

$I(\mu) = 0.125N$ and $0.9\,I(\mu) = 0.1125N$.

Thus when the interval [-4, 4] is reasonably covered we expect the mean of $I_d(\mu)/N$ to exceed 0.1125. The table suggests that about 15 points are needed for suitable coverage. This design criterion seems to require a lot more design points than other design criteria.

TABLE 2

I(μ)/n Over μ = -1(0.2) 1.

K	Mean	S.D	Coefficient of Variation
2	0.0209	0.0031	0.146
3	0.0898	0.0046	0.051
4	0.0926	0.0009	0.010
5	0.0986	0.0004	0.004
6	0.1023	0.0003	0.003
7	0.1049	0.0004	0.003
8	0.1068	0.0004	0.004
9	0.1083	0.0005	0.004
10	0.1095	0.0005	0.004
11	0.1104	0.0005	0.005
12	0.1112	0.0005	0.005
13	0.1119	0.0006	0.005
14	0.1124	0.0006	0.005
15	0.1129	0.0006	0.005
16	0.1133	0.0006	0.005
20	0.1146	0.0007	0.006
30	0.1163	0.0007	0.006
50	0.1176	0.0008	0.007
100	0.1186	0.0008	0.007

The Two Parameter Model

Single parameter models are much less used than those with two parameters. The standard model used is

$$p(x) = F(\beta(x-\mu)) \qquad (5.1)$$

The parametrization $F(\alpha + \beta x)$ is equivalent to (5.1) with $\alpha = -\mu/\beta$ and the following results apply equally to both.

For a k-point design $\{x_1, x_2, \ldots, x_k\}$, the information matrix for μ and $\log\beta$ is

$$I_d(\mu,\log\beta) = \beta^2\begin{bmatrix}\sum n_i\omega_i & -\sum n_i(x_i-\mu)\omega_i \\ -\sum n_i(x_i-\mu)\omega_i & \sum n_i(x_i-\mu)^2 w_i\end{bmatrix} \qquad (5.2)$$

where n_i subjects are tested at x_i and where

$$\omega_i = \omega(\beta(x_i-\mu)) = \frac{\{f(\beta(x_i-\mu))\}^2}{F(\beta(x_i-\mu))\{1-F(\beta(x_i-\mu))\}}.$$

Taking limits as $k \to \infty$, the information matrix becomes

$$I(\mu,\log\beta) = \beta^2 N \begin{bmatrix} \int m(x)w(\beta(x-\mu))dx & -\int (x-\mu)m(x)w(\beta(x-\mu))dx \\ -\int (x-\mu)m(x)w(\beta(x-\mu))dx & \int (x-\mu)^2 m(x)w(\beta(x-\mu))dx \end{bmatrix} \tag{5.3}$$

Looking at the parameters individually we would like the asymptotic variances of their maximum likelihood estimators to be independent of the parameters. No choice of $m(x)$ will achieve that. Note that the requirement of constant asymptotic variances means that the following two integral equations should be simultaneously satisfied.

$$\frac{\beta^2 \int_{-\infty}^{\infty} m((x)(x-\mu)^2 \omega(\beta(x-\mu))dx}{\det I(\mu,\log\beta)} = k_1 \tag{5.4}$$

$$\frac{\beta^2 \int_{-\infty}^{\infty} m(x)\omega(\beta(x-\mu))dx}{\det I(\mu,\log\beta)} = k_2 \tag{5.5}$$

where k_1, k_2 are constants.

We seek $m(x)$ that satisfy (5.4) and (5.5) simultaneously. No solution has been found, and we believe that none exists.

Equations (5.4) and (5.5) imply that

$$\int_{-\infty}^{\infty} (x-\mu)^2 m(x)\omega(\beta(x-\mu))dx = K\int_{-\infty}^{\infty} m(x)\omega(\beta(x-\mu))dx \tag{5.6}$$

where $K = k_1 / k_2$. Hence we have

$$\int_{-\infty}^{\infty} m(x)\{(x-\mu)^2 - K\}\omega(\beta(x-\mu))dx = 0 \quad for\ all\ \mu,\beta \tag{5.7}$$

Now if $\{\omega(\beta(x-\mu)): -\infty<\mu<\infty,\ -\infty<x<\infty, \beta>0\}$ is proportional to a density function of a complete probability distribution, then $m(x)$ must be zero except perhaps at $x = \mu \pm \sqrt{K}$. Hence when $\omega(.)$ is proportional to a density of a complete probability distribution, no design measure would make the asymptotic variances of $\hat{\mu}$ and $\hat{log}\beta$ constants.

Now consider the determinant of the information matrix (5.3).

$$\det I(\mu,\log\beta) = \beta^4 N^2 \left\{ \int_{-\infty}^{\infty} m(x)\omega(\beta(x-\mu))dx \int_{-\infty}^{\infty} (x-\mu)^2 m(x) \omega(\beta(x-\mu))dx - \left(\int_{-\infty}^{\infty} (x-\mu)m(x)\omega(\beta(x-\mu))dx \right)^2 \right\} \tag{5.8}$$

Substitute $t = \beta(x-\mu)$ and take $m(x) = A$, hence

$$\det I(\mu,\log\beta) = N^2 A^2 \left\{ \int_{-\infty}^{\infty} \omega(t)dt \int_{-\infty}^{\infty} t^2\omega(t)dt - \left(\int_{-\infty}^{\infty} t\omega(t)dt \right)^2 \right\} \tag{5.9}$$

Thus the determinant of the information matrix is independent of the parameters when the design measure is uniform. Abdelbasit and Plackett (1983) suggested calling such designs *D*-reliable, by analogy with *D*-optimal designs. Two open questions that merit further investigation, and are not addressed here, are

[1] Do *D*-reliable designs exist when the model has more than two parameters?
[2] Can we find designs that make other functions of the information matrix constant? In other words, by analogy with *D*-reliability, do *A*-reliable and /or *E*-reliable designs exist for models with two or more parameters? We hope to investigate these problems in future work.

Going back to *D*-reliable designs, let us now consider *k*-point discrete versions of the *D*-reliable continuous design obtained above.

We can use arguments similar to those used in sections 3 and 4 for single parameters. Call $\det I(\mu, \log\beta)$ in (5.9) the total determinant. Define $t(\epsilon)$ by

$$\left(\int_{-t(\epsilon)}^{t(\epsilon)} \omega(t)dt\right)\left(\int_{-t(\epsilon)}^{t(\epsilon)} t^2\omega(t)dt\right) - \left(\int_{-t(\epsilon)}^{t(\epsilon)} t\omega(t)dt\right)^2 = \epsilon \det I(\mu,\log\beta). \tag{5.10}$$

Suppose it is known that

$$\mu_L \le \mu \le \mu_H, \quad \beta_L \le \beta \le \beta_H \tag{5.11}$$

Define

$$x_L = \mu_L - \frac{t(\epsilon)}{\beta_L}, \quad x_H = \mu_H + \frac{t(\epsilon)}{\beta_L} \qquad (5.12)$$

Thus for

$$m(x) = \begin{cases} A & x_L \le x \le x_H \\ 0 & elsewhere \end{cases} \qquad (5.13)$$

we get at least a proportion (1 -2ϵ) of the total determinant det $I(\mu, \log\beta)$. A k-point approximation to (5.13) has samples of equal sizes at points $\{x_i\}$ equally spaced between $x_1 = x_L$ and $x_k = x_H$. Again we take $A = (x_k - x_1)^{-1}$ in (5.9) for assessing the performance of k-point approximations. Issues of stability of the determinant and the proportion of the total determinant achieved by k-point designs are again explored using an example.

EXAMPLE 3: Brown (1966) considers a probit model with $0 \le \mu \le 2$, $0.1 \le \sigma \le 0.25$. In our notation these translate to

$$0 \le \mu \le 2, \quad 4 \le \beta \le 10 \qquad (5.14)$$

Take $F(.)$ logistic, then (5.10) gives

$$\left(\int_{-t(\epsilon)}^{t(\epsilon)} \omega(t)dt\right)\left(\int_{-t(\epsilon)}^{t(\epsilon)} t^2\omega(t)dt\right) = \epsilon \det I(\mu, \log\beta) \qquad (5.15)$$

Equation (5.15) holds for any distribution which is symmetric about zero. These finite integrals cannot be obtained analytically for either the logit or probit models. An alternative model, which we introduce in the following section, is used. From the results given in section 6, we expect the results obtained using this model to be equivalent to those obtained using a probit or a logit model.

For this model $\det I(\mu, \log\beta) = N^2A^2\pi^3/8$ and $t(0.9) = 5.015$. Thus (5.12) and (5.14) give

$$x_L = -1.25, \quad x_H = 3.25 \qquad (5.16)$$

Thus taking $A^{-1} = 3.25 - (-1.25) = 4.5$ in (5.9), we get

$$\det I(\mu, \log\beta) = N^2\pi^3/8 \times (4.5)^2 = 0.191 N^2$$

and a proportion of 0.9 of that is $0.172\,N^2$.

TABLE 3

The determinant of the information matrix over a grid G Uniform design measure on [-1, 25, 3.25].

K	Mean	S.D	Coefficient of Variation
4	0.0173	0.0320	1.08488
5	0.0531	0.0536	1.0097
6	0.0919	0.0550	0.5984
7	0.1134	0.0480	0.4236
8	0.1306	0.0412	0.3152
9	0.1474	0.0283	0.1922
10	0.1551	0.0209	0.1345
11	0.1574	0.0115	0.0729
12	0.1602	0.0097	0.0608
13	0.1631	0.0042	0.0256
14	0.1650	0.0016	0.0099
15	0.1666	0.00094	0.0057
16	0.1681	0.00067	0.0040
17	0.1694	0.00067	0.0040
18	0.1706	0.00071	0.0042
19	0.1716	0.00075	0.0044
20	0.1725	0.00078	0.0045
30	0.1786	0.00104	0.0058
50	0.1835	0.00127	0.0069
100	0.1872	0.00146	0.0078

For a k - point design, with equal sample sizes, the determinant of the information matrix in (5.2) is

$$\det I_d(\mu, \log\beta) = \frac{N^2\beta^2}{k^2}\left\{\sum_{i=1}^{k} n_i\omega_i \sum_{i=1}^{k} n_i(x_i-\mu)^2\omega_i - \left(\sum_{i=1}^{k} n_i(x_i-\mu)\omega_i\right)^2\right\} \qquad (5.17)$$

We thus expect det $I_d(\mu, \log\beta)/N^2$ to exceed 0.172 for a reasonable coverage of the interval [-1.25, 3.25]. The determinant in (5.17) is computed for equally spaced design between -1.25 and 3.25 over a grid $U \times B$ where $U = 0(0.2)\ 2$ and B = 4(1)10. The means, standard deviations and the coefficients of variation of N^{-2} det I_d (μ, log β) are presented in Table 3 for different values of k. The table shows that

(1) Minimum variability is achieved with about 16 points. Increases in the coefficient of variation for $k > 17$ are very small. The determinant looks fairly stable with 12 points or more.

(2) About 19 points are needed for reasonable coverage of the interval [-1.25, 3.25]. As noted before these designs seem to suggest many more points than the number suggested by other criteria.

An Alternative Model

Calculations for Example 3 above were not possible for standard models, because the weight functions $\omega(t)$ resulting from these models make integrals hard to evaluate. We can specify a convenient weight function $\omega(t)$ and find the corresponding model (Abdelbasit and Plackett 1983). Note that from (3.3)

$$\omega(t) = (f(t))^2/F(t)\{1-F(t)\}$$

let $h(t)$ be an arbitrary density function. A general weight function involving $h(t)$ is

$$\omega(t) = \pi^2 a^2 h^2(at) \qquad (6.1)$$

corresponding to the model

$$p(t) = F(t) = \sin^2\left\{\frac{\pi}{2} H(at)\right\} \qquad (6.2)$$

where $H(t)$ is the distribution function corresponding to the density $h(t)$ and a is any positive constant. This gives a general model, different forms of which may be used by choosing different functions $h(t)$.

Taking $h(t)$ the normal density leads to

$$\omega(t) = \frac{\pi a^2}{2} \exp(-a^2 t^2) \qquad (6.3)$$

and

$$F(t) = \sin^2\left\{\frac{\pi}{2} \Phi(at)\right\} \qquad (6.4)$$

where Φ is the cumulative distribution function of the standard normal. Since a is quite arbitrary, a sensible choice for it is the one that normalizes $\omega(t)$, i.e. make $\int \omega(t)\,dt = 1$. This gives $a = \frac{2}{\pi\sqrt{\pi}}$ and hence

$$\omega(t) = \frac{2}{\pi^2} \exp(-4t^2/\pi^3) \qquad (6.5)$$

$$F(t) = \sin^2\left\{\frac{\pi}{2} \Phi\left(\frac{2t}{\pi\sqrt{\pi}}\right)\right\} \qquad (6.6)$$

The model (6.6) is the one used for the calculations in Example 3 above.

TABLE 4

Comparison of the probit, logit models and our model of equation (6.6).

x	Logit	(6.6)	Probit
0	0.50000	0.50000	0.50000
0.2	0.57912	0.57924	0.57926
0.4	0.65437	0.65526	0.65542
0.6	0.72261	0.72526	0.72574
0.8	0.78187	0.78715	0.78814
1.0	0.83143	0.83976	0.84134
1.2	0.87157	0.88279	0.88493
1.4	0.90327	0.91670	0.91924
1.6	0.92779	0.94247	0.94520
1.8	0.94646	0.96139	0.96407
2.0	0.96052	0.97481	0.97725
2.2	0.97099	0.98403	0.98610
2.4	0.97875	0.99016	0.99180
2.6	0.98447	0.99410	0.99534
2.8	0.98866	0.99657	0.99745
3.0	0.99713	0.99806	0.99865
3.5	0.99626	0.99959	0.99977
4.0	0.99831	0.99993	0.99997
4.5	0.99924	0.99999	1.00000
5.0	0.99966	1.00000	1.00000

A question of interest is how does the model given by (6.6) compare with the standard models, specially the probit and logit models. Notice that only the shape parameters of the three curves are of interest, so we may neglect the location parameters. We therefore compare the functions

(i) Logit model $p(x) = (1 + \exp(-\alpha x))^{-1}$

(ii) Probit model $p(x) = \Phi(\beta x)$

(iii) Our model $p(x) \;=\; \sin^2\left\{\frac{\pi}{2}\Phi(\gamma x)\right\}$

Cox and Snell (1989) compared (i) and (ii) with other models by equating them at the 80% point. Finney (1978, pp 362-4) compared the same models by standardizing them so that they all have zero mean and unit variance. The probit and logit models are known to agree closely except in the extreme tails.

We note that all three functions above have the same value 1/2 at $x = 0$. We avoid imposing any constraints and simply equate three term Taylor expansions of these functions about zero. Setting one of the three parameters α, β, γ equal to 1, the other two are obtained. Note further that all three distributions are symmetric about zero and the second derivative vanishes at zero for all three functions (i), (ii) and (iii) above. Thus equating three term Taylor expansions reduces to equating the three densities at $x = 0$ and only positive values of x are needed for the comparison. Setting $\beta = 1$, we get $\alpha = 2\sqrt{2}/\sqrt{\pi}$ and $\gamma = 2/\pi$. With these values of α, β, and γ, the values of the three functions above were calculated for different values of x. The results are shown in Table 4. The table shows that our model is very close to the probit and fairly close to the logit. Actually the curve for our model lies entirely between the curves for the probit and logit, with logit the lower and probit the higher of the curves.

References

ABDELBASIT, K.M. and PLACKETT, R.L. 1981. Experimental design for categorical data. International Statistical Reviews, **49**, 111-126.

ABDELBASIT, K.M. and PLACKETT, R.L. 1982. Experimental design for joint action, Biometrics, **38**, 171-179.

ABDELBASIT, K.M. and PLACKETT, R.L. 1983. Experimental design for Binary data. Journal of the American Statistical Association, **78**, 90-98.

BROWN, W.W. 1966. Planning of Quantal assay of potency. Biometrics **22**, 322-329.

CHALONER, K. and LARNTZ, K. 1989. Optimal Bayesian designs applied to logistic regression experiments. Journal of Statistical Planning and Inference, **21**, 191-208.

COCHRAN, W.G. 1973. Experiments for Non-linear Functions. Journal of the American Statistical Association, **68**, 771-781.

COX, D.R. 1988. A note on design when response has an exponential family distribution. Biometrika, **75**, 161-164.

COX, D.R. and Snell E.J. 1989. *The Analysis of Binary Data.* London, Chapman & Hall.

FINNEY, D.J. 1978. *Statistical Methods in Biological Assays.* 3rd ed, London: Griffin.

FISHER, R.A. 1922. On the Mathematical Foundation of Theoretical Statistics. Philosophical transaction of the Royal Society of London, Ser. A. **222**, 309-368.

FISHER, R.A. 1966. *The design of experiment*, 8th ed., Edinburgh: Oliver & Boyd.

FORD, I. TITTERINGTON, D.M. and KITSOS, C. 1989. Recent Advances in Non-linear Experimental design. Technometrics, **31**, 49-60.

KHAN, M.K. and YAZIDI, A.A. 1988. On D-optimal designs for binary data. Journal of Statistical Planning and Inference, **18**, 83-91.

MINKIN, S. 1987. Optimal design for binary data. Journal of the American Statistical Association, **82**, 1098-1103.

PLACKETT, R.L. 1981. The Analysis of Categorical data. 2nd ed., London, Griffin.

SILVEY, S.D. 1980. Optimal design. London, Chapman & Hall.

SITTER, R.R. 1992. Robust designs for binary data. Biometrics, **48**, 1145 - 1155.

WU, C.F. JEFF. 1985. Efficient sequential designs with binary data. Journal of the American Statistical Association, **80**, 974-984.

Received 17 February 1997
Accepted 7 July 1998