Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 4059–4070
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Locally D-optimal designs for multistage models and heteroscedastic polynomial regression models

Zhide Fang[a,*], Douglas P. Wiens[b], Zheyang Wu[a]

[a]*Department of Mathematics, University of New Orleans, New Orleans LA 70148, USA*
[b]*Department of Mathematical and Statistical Sciences, University of Alberta, USA*

Received 31 December 2003; accepted 19 November 2004
Available online 23 May 2005

Abstract

We consider the construction of locally D-optimal designs for a nonlinear, multistage model in which one observes a binary response variable with expected value $P(x;\boldsymbol{\theta}) = H(\theta_0 + \theta_1 x + \cdots + \theta_k x^k)$. Here H is any twice differentiable distribution function. Our results apply as well to heteroscedastic polynomial regression models, under mild conditions on the efficiency function.

MSC: Primary 62K05; 62F35; Secondary 62J05

Keywords: Bioassay; D-optimality; Dose response; Optimal design

1. Introduction

Consider the design problem for a dose-response model, in which one observes, with error, a binary response variable $Y(x)$ with expected value

$$P(x;\boldsymbol{\theta}) = H(\theta_0 + \theta_1 x + \cdots + \theta_k x^k), \tag{1}$$

corresponding to an input variable (dose level) x lying in an interval $[0, b]$, $0 < b < \infty$. Here H is any twice differentiable distribution function. The parameters θ_j are unknown, and are

* Corresponding author. Tel.: +1 504 2806108; fax: +1 504 2805516.
E-mail addresses: zfang@math.uno.edu (Z. Fang), doug.wiens@ualberta.ca (D.P. Wiens), ywu2@uno.edu (Z. Wu).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2004.11.015

to be estimated. In oncological studies, (1) with $H(z)=1-\exp(-z)$ for $z=\sum_{i=0}^{k}\theta_i x^i \geqslant 0$ is called a k-stage model, and $P(x;\boldsymbol{\theta})$ gives the lifetime probability of developing cancer after exposure at level x. The k-stage model can be adopted to assess the carcinogenic risks resulting from exposure to environmental chemicals. The model assumes that the mechanism of carcinogenesis can be interpreted as a series of k somatic mutations at the cellular level. After going through this series of mutational stages, a cell becomes malignant and proceeds to develop into a tumour. See, for example, Armitage and Doll (1954), Portier and Hoel (1983), Portier and Elder (1990) and references therein for detailed discussion of the biological background of the model.

Suppose that n subjects are tested at n, not necessarily distinct, dose levels $x_1,\ldots,x_n$. The design, denoted by ξ, is defined to be the measure placing mass n^{-1} at each x_j. Set $\boldsymbol{\theta}=(\theta_0,\ldots,\theta_k)'$ and $\mathbf{f}(x)=(1,x,\ldots,x^k)'$. The vector $\boldsymbol{\theta}$ is typically estimated by maximum likelihood. From the Bernoulli likelihood

$$Y(x_i)\sim \mathrm{bin}(1, P(x_i;\boldsymbol{\theta})=H(\boldsymbol{\theta}'\mathbf{f}(x_i))),$$

one obtains the information matrix of the design as

$$\begin{aligned} I(\boldsymbol{\theta};\xi) &= \sum_{i=1}^{n}\mathbf{f}(x_i)\lambda(x_i;\boldsymbol{\theta})\mathbf{f}'(x_i)\\ &= \int_0^b \mathbf{f}(x)\lambda(x;\boldsymbol{\theta})\mathbf{f}'(x)\xi(\mathrm{d}x), \end{aligned} \tag{2}$$

where, with $h=H'$,

$$\lambda(x;\boldsymbol{\theta})=\left\{\frac{h^2(z)}{H(z)(1-H(z))}\right\}_{|z=\boldsymbol{\theta}'\mathbf{f}(x)}. \tag{3}$$

In the theory of optimal designs, the criterion for optimality is usually defined through some scalar function of the information matrix. Then the optimal design is the one which maximizes (or minimizes) this function among a class of candidate designs. For the multi-stage model (1), the information matrix depends on the unknown parameter vector $\boldsymbol{\theta}$. Thus the design problem becomes more difficult than that for linear regression models. One way to handle this difficulty is to guess the true value, or obtain a good initial estimate, of $\boldsymbol{\theta}$, and to then construct the *locally optimal* design. Locally optimal designs are introduced and discussed by Chernoff (1953), Fedorov (1972), Silvey (1980), Ford et al. (1989) and others.

In this paper, we seek the locally D-optimal designs for model (1), i.e., designs maximizing $\log|I(\boldsymbol{\theta};\xi)|$ for a prespecified choice of $\boldsymbol{\theta}$. Although $\boldsymbol{\theta}$ is of course never known in practice, locally optimal designs are of interest, especially when good initial parameter estimates are available or when sequential designs can be carried out in batches—the estimates of parameters based on the previous batches can serve as the initial estimates in order to construct the locally optimal design for the next batch. See Ford et al. (1992) and Sitter and Fainaru (1997) for a detailed discussion.

Our model is different from those considered in Sitter and Fainaru (1997). They construct, by geometric methods, locally optimal designs for bioassay models $P(x;\boldsymbol{\theta})=H(\theta_1(x-\theta_2))$, where H is a distribution function with a symmetric density. Logistic and probit models are special cases addressed there. Design problems for an exponential regression model, under which the response at x is $H(z)$, observed with normally distributed error, are discussed in Fang and Wiens (2004).

From (2), one sees that our problem is equivalent to that of finding the D-optimal design for a heteroscedastic polynomial regression model of degree k, with efficiency function $\lambda(x;\boldsymbol{\theta})$. The efficiency function must be nonnegative, but need not be of the form (3). Our framework is thus more general than has previously been considered—see Dette et al. (2003) for instance.

To be implementable ξ must be an *exact* design, i.e., it must have atoms which are multiples of n^{-1}. Since the exact design problem often leads to an intractable integer programming problem, we restrict ourselves to *approximate* designs, that is, discrete but otherwise arbitrary probability measures on the design space. Approaches to the implementation of approximate designs are discussed in Pukelsheim (1993, Chapter 12).

2. Optimal designs

In this section we describe the D-optimal designs when the information matrix is of the form (2) for a positive efficiency function $\lambda(x;\boldsymbol{\theta})$. Special cases and applications will be considered in the next section. Without loss of generality, we take $b=1$. We define $\omega(x;\boldsymbol{\theta})=1/\lambda(x;\boldsymbol{\theta})$, with derivative $\dot{\omega}(x;\boldsymbol{\theta})=(\partial/\partial x)\omega(x;\boldsymbol{\theta})$.

For $x_k\in(0,1]$ let $\mathscr{S}(x_k)$ be the $k-1$-dimensional simplex

$$\mathscr{S}(x_k)=\{(x_1,\ldots,x_{k-1})|0<x_1\leqslant\cdots\leqslant x_{k-1}\leqslant x_k\}.$$

Let $\xi(\mathbf{x};x_k)$ be a $k+1$-point design placing equal mass at each support point $\{x_0=0,x_1,\ldots,x_{k-1},x_k\}$, with $\mathbf{x}=(x_1,\ldots,x_{k-1})\in\mathrm{Int}(\mathscr{S}(x_k))$. Then the information matrix factors as

$$I(\boldsymbol{\theta};\xi(\mathbf{x};x_k))=\frac{1}{k+1}\mathbf{F}(\mathbf{x};x_k)\boldsymbol{\Omega}^{-1}(\mathbf{x};x_k,\boldsymbol{\theta})\mathbf{F}'(\mathbf{x};x_k),$$

where $\mathbf{F}(\mathbf{x};x_k)_{(k+1)\times(k+1)}$ has columns $\mathbf{f}(x_i)$, $i=0,\ldots,k$, and $\boldsymbol{\Omega}(\mathbf{x};x_k,\boldsymbol{\theta})=\mathrm{diag}(\omega(x_0;\boldsymbol{\theta}),\ldots,\omega(x_k;\boldsymbol{\theta}))$. The log-determinant is

$$\log|I(\boldsymbol{\theta};\xi(\mathbf{x};x_k))|=\log(|\mathbf{F}(\mathbf{x};x_k)|^2)-\log|\boldsymbol{\Omega}(\mathbf{x};x_k,\boldsymbol{\theta})|-(k+1)\log(k+1). \quad (4)$$

Since $|\mathbf{F}(\mathbf{x};x_k)|\to 0$ as $(x_1,\ldots,x_{k-1})$ approaches the boundary of $\mathscr{S}(x_k)$, $\log|I(\boldsymbol{\theta};\xi(\mathbf{x};x_k))|=-\infty$ everywhere on the boundary and there is a point $\mathbf{t}=(t_1,\ldots,t_{k-1})$ in the interior of $\mathscr{S}(x_k)$ at which $\log|I(\boldsymbol{\theta};\xi(\cdot;x_k))|$ attains its maximum. Such a point is necessarily a stationary point.

To characterize the stationary point, let $\xi(\mathbf{t};x_k)$ put equal weight on the points $\{t_0=0<t_1<\cdots<t_{k-1}<t_k=x_k\}$. Let $\mathbf{F}^{-1}(\mathbf{t};x_k)$ have rows $\mathbf{a}_i'(\mathbf{t};x_k)$ for $i=0,\ldots,k$. Note that

for $j=0,\ldots,k$,

$$\begin{aligned}\frac{\partial}{\partial t_j}\log|I(\boldsymbol{\theta};\xi(\mathbf{t};x_k))| &= \frac{\partial}{\partial t_j}\log|\mathbf{F}(\mathbf{t};x_k)|^2-\frac{\dot{\omega}(t_j;\boldsymbol{\theta})}{\omega(t_j;\boldsymbol{\theta})}\\ &= 2\,\mathrm{tr}\left(\mathbf{F}^{-1}(\mathbf{t};x_k)\frac{\partial}{\partial t_j}\mathbf{F}(\mathbf{t};x_k)\right)-\frac{\dot{\omega}(t_j;\boldsymbol{\theta})}{\omega(t_j;\boldsymbol{\theta})}\\ &= 2\mathbf{a}_j'(\mathbf{t};x_k)\dot{\mathbf{f}}(t_j)-\frac{\dot{\omega}(t_j;\boldsymbol{\theta})}{\omega(t_j;\boldsymbol{\theta})}.\end{aligned}$$

Thus a stationary point $\mathbf{t}$ necessarily satisfies

$$2\mathbf{a}_j'(\mathbf{t};x_k)\dot{\mathbf{f}}(t_j)=\frac{\dot{\omega}(t_j;\boldsymbol{\theta})}{\omega(t_j;\boldsymbol{\theta})},\quad j=1,\ldots,k-1. \tag{5}$$

Recall that $t_k=x_k\in(0,1]$. With $\mathbf{t}(x_k)$ defined implicitly by (5), we now define

$$x_k=\arg\max_{x\in(0,1]}\log|I(\boldsymbol{\theta};\xi(\mathbf{t}(x);x))|.$$

We argue in Remark 3 below that t_k and $\mathbf{t}(t_k)$ are unique.

Before stating our main result, we require a preliminary lemma.

Lemma 1. *Let* $\mathbf{V}$ *be the* $(k+1)\times(k+1)$ *matrix with columns* $\mathbf{f}(x_i)$, $i=0,\ldots,k$ *where* $0\leqslant x_0<\cdots<x_k$. *Then the* $(1,2)$ *element of* $\mathbf{V}^{-1}$ *is*

$$\mathbf{V}^{12}=\left\{\prod_{j=1}^{k}\frac{x_j}{x_j-x_0}\right\}\cdot\left\{-\sum_{i=1}^{k}\frac{1}{x_i}\right\}. \tag{6}$$

Proof. Assume first that $x_0>0$. Then

$$\mathbf{V}=\begin{pmatrix}1&1&1&\cdots&1\\ x_0&x_1&x_2&\cdots&x_k\\ x_0^2&x_1^2&x_2^2&\cdots&x_k^2\\ \vdots&\vdots&\vdots&\cdots&\vdots\\ x_0^k&x_1^k&x_2^k&\cdots&x_k^k\end{pmatrix},$$

a Vandermonde matrix whose determinant is well known:

$$|\mathbf{V}|=\prod_{0\leqslant i<j\leqslant k}(x_i-x_j)=\left\{\prod_{j=1}^{k}(x_0-x_j)\right\}\cdot\prod_{1\leqslant i<j\leqslant k}(x_i-x_j). \tag{7}$$

If instead $|\mathbf{V}|$ is evaluated by expanding along the second row, then the coefficient of the *linear* term in x_0 is -1 times the cofactor of $\mathbf{V}_{21}$, i.e., is $|\mathbf{V}|\mathbf{V}^{12}$. From (7), this coefficient is

$$\left\{\sum_{i=1}^{k}\prod_{1\leqslant j\leqslant k, j\neq i}(-x_j)\right\}\prod_{1\leqslant i<j\leqslant k}(x_i-x_j)$$
$$=\left\{\prod_{j=1}^{k}(-x_j)\sum_{i=1}^{k}\frac{1}{-x_i}\right\}\prod_{1\leqslant i<j\leqslant k}(x_i-x_j).$$

Thus

$$\mathbf{V}^{12}=\frac{\left\{\prod_{j=1}^{k}(-x_j)\sum_{i=1}^{k}\frac{1}{-x_i}\right\}\prod_{1\leqslant i<j\leqslant k}(x_i-x_j)}{\prod_{0\leqslant i<j\leqslant k}(x_i-x_j)},$$

which reduces to (6). Since this is continuous in x_0 we can drop the restriction that x_0 be positive. □

Theorem 1. *Let ξ_* be the design placing equal mass at each point $\{t_0=0<t_1<\cdots<t_{k-1}<t_k\}$, with t_k and $\mathbf{t}=\mathbf{t}(t_k)$ as above. Assume that $\omega(x;\boldsymbol{\theta})$ is differentiable at least $2k+1$ times, that $\omega^{(2k+1)}(x;\boldsymbol{\theta})$ does not change sign for $x\in[0,1]$, and that $\omega(0;\boldsymbol{\theta})>0$, $\dot{\omega}(0;\boldsymbol{\theta})\geqslant 0$. Then ξ_* is the locally D-optimal design when the information matrix is given by* (2).

Proof. To show that ξ_* is optimal within the class of all designs, rather than merely within the class of $k+1$-point designs, we invoke the celebrated Equivalence Theorem (Kiefer, 1974). By this, a design ξ_* with information matrix (2) is locally D-optimal if and only if the standardized variance $d(x;\xi_*)\triangleq\lambda(x;\boldsymbol{\theta})\mathbf{f}'(x)I^{-1}(\boldsymbol{\theta},\xi_*)\mathbf{f}(x)$ is at most $k+1$, with the maximum being obtained at the support points of ξ_*. Equivalently, we are to show that

$$m(x;t_k)\leqslant 0 \quad \text{for all } x\in[0,1], \tag{8}$$

where

$$\mathbf{g}(x)\triangleq\mathbf{F}^{-1}(\mathbf{t};t_k)\mathbf{f}(x),$$
$$m(x;t_k)\triangleq\mathbf{g}'(x)\boldsymbol{\Omega}(\boldsymbol{\theta},\xi_*)\mathbf{g}(x)-\omega(x;\boldsymbol{\theta})$$
$$=\mathbf{f}'(x)\left[\sum_{i=0}^{k}\mathbf{a}_i(\mathbf{t};t_k)\omega(t_i;\boldsymbol{\theta})\mathbf{a}_i'(\mathbf{t};t_k)\right]\mathbf{f}(x)-\omega(x;\boldsymbol{\theta}).$$

Note that $\mathbf{g}(t_i)=(0,\ldots,0,1,0,\ldots,0)'$, with '1' in the $(i+1)$th spot for $i=0,\ldots,k$. Thus we have equality in (8) when x is a support point of ξ_*.

We calculate that

$$\dot{m}(x;t_k)=2\mathbf{f}'(x)\left[\sum_{i=0}^{k}\mathbf{a}_i(\mathbf{t};t_k)\omega(t_i;\boldsymbol{\theta})\mathbf{a}_i'(\mathbf{t};t_k)\right]\dot{\mathbf{f}}(x)-\dot{\omega}(x;\boldsymbol{\theta}).$$

For $j=0,\dots,k$, we have $\mathbf{f}'(t_j)\mathbf{a}_i(\mathbf{t};t_k)=\delta_{ij}$ (Kronecker's delta). Thus for $j=1,\dots,k-1$,

$$\begin{aligned}\dot{m}(t_j;t_k)&=2\omega(t_j;\boldsymbol{\theta})\mathbf{a}_j'(\mathbf{t};t_k)\dot{\mathbf{f}}(t_j)-\dot{\omega}(t_j;\boldsymbol{\theta})\\&=0,\end{aligned}$$

by (5). For $j=0$, using Lemma 1,

$$\begin{aligned}\dot{m}(0;t_k)&=2\omega(0;\boldsymbol{\theta})\mathbf{a}_0'(\mathbf{t};t_k)\dot{\mathbf{f}}(0)-\dot{\omega}(0;\boldsymbol{\theta})\\&=2\omega(0;\boldsymbol{\theta})\mathbf{F}^{-1}(\mathbf{t})_{12}-\dot{\omega}(0;\boldsymbol{\theta})\\&=-2\omega(0;\boldsymbol{\theta})\sum_{i=1}^{k}\frac{1}{t_i}-\dot{\omega}(0;\boldsymbol{\theta})\\&<0.\end{aligned}$$

Finally, for $j=k$,

$$\dot{m}(t_k;t_k)=2\omega(t_k;\boldsymbol{\theta})\mathbf{a}_k'(\mathbf{t};t_k)\dot{\mathbf{f}}(t_k)-\dot{\omega}(t_k;\boldsymbol{\theta}).$$

If $t_k<1$ then it is a critical point, hence $\dot{m}(t_k;t_k)=0$. Otherwise, $\log|I(\boldsymbol{\theta};\xi(\mathbf{t}(x);x))|$ is increasing at $x=1$, so that $\dot{m}(t_k;t_k)=\dot{m}(1;1)\geqslant 0$.

Now note that $m(x;t_k)$ is a polynomial of degree $2k-\omega(x;\boldsymbol{\theta})$. Thus the $(2k+1)$th derivative of $m(x;t_k)$ does not change sign, so that $\ddot{m}(x;t_k)$ has at most $2k-1$ zeros and $\dot{m}(x;t_k)$ has at most $2k$ zeros in [0, 1] (including multiplicities). It now follows that $m(x;t_k)$ cannot be >0 anywhere in [0, 1]—otherwise $\ddot{m}(x;t_k)$, will have more than $2k-1$ zeros if $m(x;t_k)$ changes sign at some t_i, or $\dot{m}(x;t_k)$ will have more than $2k$ zeros if $m(x;t_k)$ does not change sign at any t_i. □

Remarks.

1. The importance of Theorem 1 is that it shows that the design which places equal masses at each of 0 and the coordinates of the point $(t_1,\dots,t_k)$ maximizing the determinant of the corresponding information matrix is locally D-optimal within the class of *all* designs. The very general form of $\omega(x;\boldsymbol{\theta})$ precludes a complete determination of $(t_1,\dots,t_k)$. This is explored in the special cases of the next section.Theorem 1
2. holds for the k-stage model with $H(z)=1-\exp(-z)$ $(z\geqslant 0)$ and $\theta_0,\dots,\theta_k>0$. Then $\omega(x;\boldsymbol{\theta})=\exp(\boldsymbol{\theta}'\mathbf{f}(x))-1>0$ and $\dot{\omega}(x;\boldsymbol{\theta})=(\omega(x;\boldsymbol{\theta})+1)\boldsymbol{\theta}'\dot{\mathbf{f}}(x)>0$. If $\theta_0=0$, $\omega(0;\boldsymbol{\theta})$ vanishes and Theorem 1 fails—indeed, the form of the information matrix in (2), with $\mathbf{f}(x)=(x,x^2,\dots,x^k)'$, implies that the locally D-optimal design must exclude zero. A separate but parallel development is possible, however. Let $\tilde{\xi}(\mathbf{x};x_k)$ be a k-point design placing equal mass at each support point $\{0<x_1,\dots,x_{k-1},x_k\}$, where $\mathbf{x}=(x_1,\dots,x_{k-1})\in\mathrm{Int}(\mathscr{S}(x_k))$. Let $\tilde{\mathbf{F}}(\mathbf{x};x_k)=(\mathbf{f}(x_1),\dots,\mathbf{f}(x_k))$ and $\tilde{\Omega}(\mathbf{x};x_k,\theta)=\mathrm{diag}(w(x_1;\theta),\dots,w(x_k;\theta))$.Then

$$I(\boldsymbol{\theta};\tilde{\xi}(\mathbf{x};x_k))=\frac{1}{k}\tilde{\mathbf{F}}(\mathbf{x};x_k)\tilde{\boldsymbol{\Omega}}^{-1}(\mathbf{x};x_k,\boldsymbol{\theta})\tilde{\mathbf{F}}'(\mathbf{x};x_k).$$

Replace $\mathbf{F}(\mathbf{x};x_k)$, $\Omega(\mathbf{x};x_k,\theta)$ and $I(\theta;\xi(\mathbf{x};x_k))$ by $\tilde{\mathbf{F}}(\mathbf{x};x_k)$, $\tilde{\Omega}(\mathbf{x};x_k,\theta)$ and $I(\boldsymbol{\theta};\tilde{\xi}(\mathbf{x};x_k))$, respectively, and define t_k and $\mathbf{t}(t_k)$ as above. Assume that $\omega(x;\boldsymbol{\theta})$ is differentiable at least $2k+1$ times, that $\omega^{(2k+1)}(x;\boldsymbol{\theta})$ does not change sign for $x \in [0,1]$, and that $\omega(0;\boldsymbol{\theta})=0$, $\dot{\omega}(0;\boldsymbol{\theta})>0$. Then, by L'Hospital's rule, we have $d(0,\tilde{\xi})=0$. Thus $m(0,t_k)=0$. By the fact that $\dot{m}(0,t_k)=-\dot{\omega}(0;\boldsymbol{\theta})<0$, and arguments similar to those used in the proof of Theorem 1, we conclude that $\tilde{\xi}_*(\mathbf{t};t_k)$ is locally D-optimal in the class of all designs.

3. Uniqueness of t_k and $\mathbf{t}(t_k)$ in Theorem 1 follows from the uniqueness of the D-optimal design. If two designs $\xi_1 : \{0, x_1, \ldots, x_k\}$ and $\xi_2 : \{0, y_1, \ldots, y_k\}$ are locally D-optimal, then $\log|I(\theta,\xi_1)| = \log|I(\theta,\xi_2)|$. Then the design $\xi_0=(\xi_1+\xi_2)/2$ is also locally D-optimal because $\log|I(\boldsymbol{\theta};\xi_0)|$ is concave:

$$\log|I(\boldsymbol{\theta};\xi_0)| \geqslant (\log|I(\boldsymbol{\theta};\xi_1)| + \log|I(\boldsymbol{\theta};\xi_2)|)/2 = \log|I(\theta,\xi_1)|.$$

This implies that

$$m_0(x) \triangleq \mathbf{f}'(x)I^{-1}(\boldsymbol{\theta},\xi_0)\mathbf{f}(x) - (k+1)\omega(x;\boldsymbol{\theta}) \leqslant 0,$$

with equality holding at the support points of ξ_0. Thus $\dot{m}_0(x)$ has at least $2k+1$ zeros, contradicting the fact that $\omega^{(2k+1)}(x;\boldsymbol{\theta})$ does not change sign for $x \in [0,1]$.

3. One- and two-stage models

As pointed out by Guess et al. (1977) and Land (1980), when applying multistage models to study the potential risk resulting from exposure to environmental chemicals, the number of affected stages is typically at most two. Thus, in this section, we present the optimal designs for (1) when $H(z)=1-\exp(-z)$ $(z\geqslant 0)$ and $k=1$ or 2. Specifically, we obtain optimal designs separately for the following models:

$$\text{Model I: } P(x;\boldsymbol{\theta}) = 1-\mathrm{e}^{-\theta_1 x}, \quad x\in[0,1], \quad \theta_1>0,$$
$$\text{Model II: } P(x;\boldsymbol{\theta}) = 1-\mathrm{e}^{-(\theta_0+\theta_1 x)}, \quad x\in[0,1], \quad \theta_0,\theta_1>0,$$
$$\text{Model III: } P(x;\boldsymbol{\theta}) = 1-\mathrm{e}^{-(\theta_0+\theta_2 x^2)}, \quad x\in[0,1], \quad \theta_0,\theta_2>0,$$
$$\text{Model IV: } P(x;\boldsymbol{\theta}) = 1-\mathrm{e}^{-(\theta_0+\theta_1 x+\theta_2 x^2)}, \quad x\in[0,1], \quad \theta_0,\theta_1,\theta_2>0,$$
$$\text{Model V: } P(x;\boldsymbol{\theta}) = 1-\mathrm{e}^{-(\theta_1 x+\theta_2 x^2)}, \quad x\in[0,1], \quad \theta_1,\theta_2>0.$$

3.1. Models I, II and III

For Model II Theorem 1 applies, and shows that the D-optimal design places masses of $1/2$ at each of 0 and x_1, where x_1 maximizes

$$s(x) \triangleq 2\log x - \log\omega(x;\boldsymbol{\theta}).$$

For any fixed $\theta_0 \geqslant 0$, denote by $\theta^*(\theta_0) \in (1,2)$ the unique solution of the equation

$$r(\theta;\theta_0) \triangleq (2-\theta)\mathrm{e}^{\theta_0+\theta} - 2 = 0.$$

Table 1
Locally D-optimal designs for Models II and III

θ_0	θ^*	θ_1	x_1	θ_2	x_2
.5	1.799	2.0	.8997	2	.9485
		2.5	.7197	3	.7745
		3.0	.5998	4	.6707
2	1.962	2.0	.9810	2	.9904
		2.5	.7849	3	.8087
		5.0	.3924	10	.4429

The properties of θ^* come from the observations that $r(\theta;\theta_0)$ is a concave function of $\theta\geqslant 0$, with $0 < r(0;\theta_0) < r(1;\theta_0)$ and $r(2;\theta_0) < 0$. As $\theta_0 \to \infty$ we have that $\theta^* \to 2$. We calculate that

$$s'(x) = \frac{r(\theta_1 x;\theta_0)}{x\omega(x;\boldsymbol{\theta})},$$

from which we obtain part (i) of the following Corollary. Part (ii) is then immediate, upon identifying the regressor x^2 of Model III with x of Model II.

Corollary 1. (i) *For Model II the locally D-optimal design puts equal mass on each of* $\{0, 1\}$ *if* $\theta_1 < \theta^*(\theta_0)$, *and on each of* $\{0, x_1\}$ *if* $\theta_1 \geqslant \theta^*(\theta_0)$, *where* x_1 *is the zero of* $r(\theta_1 x;\theta_0)$.

(ii) *For Model III the locally D-optimal design puts equal mass on each of* $\{0, 1\}$ *if* $\theta_2 < \theta^*(\theta_0)$, *and on each of* $\{0, x_1\}$ *if* $\theta_2 \geqslant \theta^*(\theta_0)$, *where* x_1 *is the zero of* $r(\theta_2 x^2;\theta_0)$.

We present some numerical values of θ^*, x_1, x_2, for two values of θ_0, in Table 1. We observe that if θ_1 (or θ_2) is large, the designs differ markedly from the classical D-optimal designs for the linear regression model with expected response $\log(1 - P(x;\boldsymbol{\theta}))$.

Note that Model I is Model II with $\theta_0 = 0$. Remark 2 of Section 2 applies, and gives

Corollary 2. *For Model I the locally D-optimal design puts all mass on* $\{1\}$ *if* $\theta_1 < \theta^*(0)$, *and on* $\{x_1\}$ *if* $\theta_1 \geqslant \theta^*(0)$, *where* x_1 *is the zero of* $r(\theta_1 x; 0)$.

3.2. Models IV and V

Theorem 1 applies to Model IV. Our development hinges on the properties of two functions

$$t(x_1, x_2;\boldsymbol{\theta}) = 2(\log x_1 + \log x_2 + \log(x_2 - x_1)) - (\log\omega(x_1;\boldsymbol{\theta}) + \log\omega(x_2;\boldsymbol{\theta})),\quad 0\leqslant x_1\leqslant x_2\leqslant 1,$$

$$u(x_1;\boldsymbol{\theta}) = x_1(1 - x_1)\dot{\omega}(x_1;\boldsymbol{\theta}) - 2(1 - 2x_1)\omega(x_1;\boldsymbol{\theta}),\quad 0\leqslant x_1\leqslant .5.$$

The second derivative of $u(x_1;\boldsymbol{\theta})$ is

$$\ddot{u}(x_1;\boldsymbol{\theta}) = 6\dot{\omega}(x_1;\boldsymbol{\theta}) + x_1(1 - x_1)\dddot{\omega}(x_1;\boldsymbol{\theta}) > 0,$$

so that $u(x_1;\boldsymbol{\theta})$ is a convex function, negative at $x_1 = 0$ and positive at $x_1 = .5$. Define $x_*(\boldsymbol{\theta})$ to be the unique zero of $u(x_1;\boldsymbol{\theta})$ in $(0, .5)$.

Corollary 3. *For Model IV, for each $x_1 \in (0, x_*(\boldsymbol{\theta})]$ define*

$$x_2 = x_1 \left(1 + \frac{2\omega(x_1; \boldsymbol{\theta})}{2\omega(x_1; \boldsymbol{\theta}) - x_1\dot{\omega}(x_1; \boldsymbol{\theta})}\right). \quad (9)$$

Then $0 < x_1 < x_2 \leqslant 1$, and for each x_2, $t(x_1, x_2; \boldsymbol{\theta})$ is maximized for fixed x_2 by x_1. If x_1 is chosen to maximize $t(x_1, x_2(x_1); \boldsymbol{\theta})$, then the design placing mass $1/3$ at each of $\{0, x_1, x_2\}$ is locally D-optimal for Model IV.

Proof. By Theorem 1 the optimal design places mass $1/3$ at each of $\{0, x_1, x_2\}$, where $0 < x_1 < x_2 \leqslant 1$. Apart from some inessential constants, the log-determinant of the information matrix is given by $t(x_1, x_2; \boldsymbol{\theta})$. Then $x_1 = x_1(x_2)$ must maximize $t(x_1, x_2; \boldsymbol{\theta})$ for fixed $x_2 \leqslant 1$, and x_2 must maximize $t(x_1(x_2), x_2; \boldsymbol{\theta})$. We calculate that

$$\frac{\partial t}{\partial x_1} = \frac{q(x_1, x_2)}{x_1(x_2 - x_1)\omega(x_1; \boldsymbol{\theta})},$$

where

$$q(x_1, x_2) = x_2(2\omega(x_1; \boldsymbol{\theta}) - x_1\dot{\omega}(x_1; \boldsymbol{\theta})) - x_1(4\omega(x_1; \boldsymbol{\theta}) - x_1\dot{\omega}(x_1; \boldsymbol{\theta})).$$

Thus x_1 satisfies $q(x_1, x_2) = 0$; this becomes (9).

It is now convenient to view (9) as defining x_2 as a function of x_1. The constraint $x_2 \leqslant 1$ becomes

$$u(x_1; \boldsymbol{\theta}) \leqslant 0, \quad (10)$$

which implies as well that $x_1\dot{\omega}(x_1; \boldsymbol{\theta}) < 2\omega(x_1; \boldsymbol{\theta})$ and $x_1 < .5$. Thus (10) holds iff $x_1 \in [0, x_*(\boldsymbol{\theta})]$. It remains only to determine x_2 maximizing $t(x_1(x_2), x_2; \boldsymbol{\theta})$. Equivalently, we are to maximize $t(x_1, x_2(x_1); \boldsymbol{\theta})$ as x_1 varies over $(0, x_*(\boldsymbol{\theta})]$, with $x_2(x_1)$ defined by (9). □

Remark.

4. A sufficient condition implying that $x_2 < 1$ in Corollary 3 is

$$2 + \frac{2}{1 - x_*(\boldsymbol{\theta})} - \frac{\dot{\omega}}{\omega}(1; \boldsymbol{\theta}) < 0. \quad (11)$$

   This is because

$$\frac{\mathrm{d}t(x_1(x_2), x_2; \boldsymbol{\theta})}{\mathrm{d}x_2}_{|x_2=1} = \left(\frac{\partial t}{\partial x_1}\frac{\partial x_1}{\partial x_2} + \frac{\partial t}{\partial x_2}\right)_{|x_2=1} = \frac{\partial t}{\partial x_2}_{|x_1=x_*(\boldsymbol{\theta}), x_2=1}$$
$$= 2 + \frac{2}{1 - x_*(\boldsymbol{\theta})} - \frac{\dot{\omega}}{\omega}(1; \boldsymbol{\theta}),$$

   since $\partial t/\partial x_1 = 0$. Thus if (11) holds then the maximizer of $t(x_1(x_2), x_2; \boldsymbol{\theta})$ must be < 1. Numerical investigations support the conjecture that (11) is necessary as well.

Table 2
Locally D-optimal designs for Model IV

θ_0	θ_1	θ_2	x_*	x_1	x_2
1	1	1	.3698	.3698	1.0000
1	1	3	.3089	.2813	.8267
1	3	1	.2867	.2823	.9689
1	3	3	.2520	.2000	.8000
3	1	1	.3908	.3908	1.0000
3	1	3	.3306	.3030	.8432
3	3	1	.3051	.3026	.9834
3	3	3	.2690	.2000	.8000

See Table 2 for some representative examples of D-optimal designs for Model IV.

Under Model V, the design must exclude zero. Note that now $\boldsymbol{\theta} = (\theta_1, \theta_2)'$.

Corollary 4. *For Model V, for each* $x_1 \in (0, x_*(\boldsymbol{\theta})]$ *define*

$$x_2 = x_1 \left(1 + \frac{2\omega(x_1; \boldsymbol{\theta})}{2\omega(x_1; \boldsymbol{\theta}) - x_1 \dot{\omega}(x_1; \boldsymbol{\theta})}\right).$$

Then $0 < x_1 < x_2 \leqslant 1$, *and* x_1 *maximizes* $t(x_1, x_2; \boldsymbol{\theta})$ *for fixed* x_2. *If* x_1 *is chosen to maximize* $t(x_1, x_2(x_1); \boldsymbol{\theta})$, *then the design placing mass* $1/2$ *at each of* $\{x_1, x_2\}$ *is locally D-optimal for Model V.*

4. Case study

The following experiment was run to test the presence, after exposure to the chemical compound *2,3,4,5 Tetra Chlorodibenzo P-Dioxin*, of internal anomalies in fetuses of rats. The data are from Table 3 of Janardan (1995) (see also Rai and Van Ryzin, 1979).

The fitted two-stage model is

$$\hat{P}(x) = 1 - \mathrm{e}^{-(0.0372x + 0.3202x^2)}.$$

The expected numbers of responses, estimated from the fitted model, are $\{0, 0.37, 0.95, 2.91, 3.01\}$. These numbers indicate that the model has a good fit to the data. Using $\theta_1 = 0.0372$, $\theta_2 = 0.3202$ as our best guesses for the model parameters, the design ξ_* of Corollary 4 is

$$\left\{ \begin{array}{ll} x_1 = 0.1664, & x_2 = 1 \\ n_1 = 68, & n_2 = 68 \end{array} \right\}.$$

This has the generalized variance $|I^{-1}(\boldsymbol{\theta}, \xi_*)| = 1.355$. The corresponding generalized variance for the design used in Janardan (1995) is 5.334, about 4 times that for the optimal design. A message from this example is that substantial savings in resources can be realized

Table 3
Toxic response data

Dose x (kg)	Animals tested	Animals with toxic response
0	24	0
.125	38	0
.25	33	1
.50	31	3
1.0	10	3

by the optimal design. The experimenter would have required only about one half as many animals to obtain the same level of accuracy, if an optimal design had been employed, and had such prior parameter estimates been available.

5. Conclusions and remarks

We have constructed locally D-optimal designs for various bioassay models, assuming that there is sufficient prior information available about the model parameters. Such information can be obtained from previous experiments, or by using a preliminary sample (training data). The obvious practical advantage of the designs so obtained is the saving of resources.

When we have only partial prior information about the parameters, for example the knowledge that they lie in a particular subset of the parameter space, the design problem can be handled in two possible ways. One is to consider the *maximin* design, which maximizes the minimum, over the possible range Θ of θ, of $\log|I(\boldsymbol{\theta};\xi)|$. Another method is to assume a prior distribution $\eta(\boldsymbol{\theta})$ on $\boldsymbol{\theta}$, and then seek a Bayesian D-optimal design, which maximizes $E_\eta[\log|I(\boldsymbol{\theta};\xi)|]$. The corresponding Equivalence Theorem states that a design ξ is Bayesian D-optimal for the k stage model if and only if

$$\int \lambda(x;\boldsymbol{\theta})\mathbf{f}'(x)I^{-1}(\boldsymbol{\theta},\xi)\mathbf{f}(x)\,\mathrm{d}\eta(\boldsymbol{\theta})\leqslant k+1$$

with the equality holding at the support points of ξ. See, for example, Dette and Wong (1998). Note that for Models I and II, if $\theta_1 < 1(<\boldsymbol{\theta}^*)$ then the locally D-optimal design places all mass in $\{0,1\}$. Since this design does then not depend on $\boldsymbol{\theta}$ it is maximin if $\{\theta_1|\boldsymbol{\theta}\in\Theta\}\subset[0,1]$, and is Bayesian D-optimal for any prior $\eta(\boldsymbol{\theta})$ for which $P_\eta(\theta_1\leqslant 1)=1$. The problems of constructing maximin designs and Bayesian D-optimal designs for other models and for other prior distributions were touched on in Fang and Wiens (2004), and are topics of further research.

Acknowledgements

The research of Douglas Wiens is supported by the Natural Sciences and Engineering Research Council of Canada. All authors are grateful for the helpful suggestions of the reviewers.

References

Armitage, P., Doll, R., 1954. The age distribution of cancer and a multistage theory of cancer. British J. Cancer 8, 1–12.

Chernoff, H., 1953. Locally optimal designs for estimating parameters. Ann. Math. Statist. 24, 586–602.

Dette, H., Wong, W.K., 1998. Bayesian D-optimal designs on a fixed number of design points for heteroscedastic polynomial models. Biometrika 84, 869–882.

Dette, H., Haines, L., Imhof, L., 2003. Bayesian and maximin optimal designs for heteroscedastic regression models. Preprint. http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm.

Fang, Z., Wiens, D.P., 2004. Bayesian minimally supported D-optimal designs for an exponential regression model. Comm. Statist. Theory Methods 33, 1187–1204.

Fedorov, V.V., 1972. Theory of Optimal Experiments. Academic Press, New York.

Ford, I., Titterington, D.M., Kitsos, C.P., 1989. Recent advances in nonlinear experimental design. Technometrics 31, 49–60.

Ford, I., Titterington, D.M., Wu, C.F.J, 1992. The use of a canonical form in the construction of locally optimal designs for nonlinear problems. J. Roy. Statist. Soc. 54, 569–583.

Guess, H., Grump, K., Peto, R., 1977. Uncertainty estimates for low-dose-rate extrapolations of animal carcinogenicity data. Cancer Research 37, 3475–3483.

Janardan, K.G., 1995. Environmental risk assessment model for low dose extrapolation. Comm. Statist. Theory Methods 24 (10), 2621–2633.

Kiefer, J., 1974. General equivalence theory for optimum designs (approximate theory). Ann. Statist. 2, 849–879.

Land, C., 1980. Estimating cancer risks from low doses of ionizing radiation. Science 209, 1197–1203.

Portier, C., Elder, L., 1990. Two-stage models of carcinogenesis, classification of agents, and design of experiments. Fund. Appl. Toxicol. 14, 444–460.

Portier, C., Hoel, D., 1983. Low-dose rate extrapolation using the multistage model. Biometrics 39, 897–906.

Pukelsheim, F., 1993. Optimal Design of Experiments. Wiley, Toronto.

Rai, K., Van Ryzin, J., 1979. Risk assessment of toxic substances using a generalized multi-hit dose response model. In: Breslow, N.E., Whittemore, A.S. (Eds.), Energy and Health Proceedings of Sims Conference Alta, SIAM, Philadelphia, June 26–30, 1978.

Silvey, S.D., 1980. Optimal Design. Chapman & Hall, London.

Sitter, R.R., Fainaru, I., 1997. Optimal designs for the logit and probit models for binary data. Canad. J. Statist. 25, 175–190.