Statistica Sinica **12**(2002), 1145-1155

OPTIMAL DESIGN OF EXPERIMENTS WITH POSSIBLY FAILING TRIALS

Lorens A. Imhof, Dale Song and Weng Kee Wong

RWTH Aachen, Scirex Corporation and University of California at Los Angeles

Abstract: We propose a method for finding optimal designs when there are potentially failing trials in the experiment. Examples are presented for polynomial models using different types of response probability functions, including situations when these response probability functions are only partially specified. Some properties of the proposed optimal designs are discussed.

Key words and phrases: Approximate designs, D-optimality, information matrix, maximin designs, missing observations.

1. Introduction

Almost all work in the field of optimal design assumes that all trials of the experiment will result in observations of the response variable. In practice, however, it is conceivable that not all responses of the trials are realized when the experiment is carried out. There are several ways, often outside the control of the experimenter, that can cause non-response. For example, in industrial experiments, instruments may be more likely to malfunction at high temperatures or pressures than under normal conditions. Missing values also tend to occur more frequently in a long term clinical trial for a slow acting disease such as rheumatoid arthritis where patients are less likely to show up for scheduled appointments as time progresses. This means that in both cases some trials do not result in observations. In these situations, designs which assume all observations are available at the end of the experiment can perform poorly. If we have prior knowledge about potentially failing trials, it is therefore desirable to incorporate such information in the design of the experiment.

To fix ideas, consider the straight-line homoscedastic regression model on the interval $[0, 1]$ and suppose that 100 observations are to be taken. The usual D-optimal design for estimating the two parameters requires that half the observations be taken at each of the end-points. When there are failing trials, fewer than 100 observations are available at the end of the experiment. If the response probabilities at 0 and 1 are different and low, the design will be unbalanced with many missing observations at $x = 0$ and $x = 1$. Consequently, the usual D-optimal design can be inefficient for estimating the model parameters.

There is very little work in the literature that addresses the design problem for experiments with potentially failing trials. An early work is Herzberg and Andrews (1976) where they proposed three criteria to assess a design in a failing trial situation. They considered a setting where experiments are allowed to have only 3 or 4 observations. Under this constraint, optimal designs were found for the simple linear and quadratic regression models. No general method was provided to construct optimal designs. Akhtar and Prescott (1986) considered central composite designs of a second order and proposed criteria robust to one or two missing observations. Recently, Hackl (1995) found exact D-optimal designs for estimating coefficients in a quadratic model when there are failing trials. He assumed the design space consists of equally spaced points and the sample size is small, with no replications allowed. The optimal design was found by comparing all possible design candidates assuming the probability of having a response is monotonic. Closed-form formulae and properties of the optimal designs were not provided because of the complexity of the problem.

The aim of this paper is to provide a general method for constructing efficient designs when there are varying probabilities of realizing responses. The method is flexible and applies to both linear and non-linear models, and different design criteria. In Section 3, we discuss D-optimal designs for polynomial models and describe how the optimal designs behave when the response probabilities change. Section 4 relaxes the earlier assumption that the response probabilities have to be completely specified. Instead it is only assumed that the response probability function belongs to a known set of plausible functions. For this situation, we propose a maximin design criterion and provide closed-form maximin D-optimal designs for a class of partially specified response probability functions. Section 5 contains a summary and a discussion of the possibility that the optimal design does not provide estimates for all the model parameters.

2. Optimality Criterion

We consider statistical models of the form

$$y(x) = f(x, \gamma) + \epsilon, \qquad x \in \mathcal{X}, \tag{2.1}$$

where $y(x)$ is the response at x and the function f is assumed known apart from the model parameters γ. The error term ϵ is normally distributed with mean zero and constant variance. The design space $\mathcal{X}$ is a given compact set; in applications, it is usually an interval. Additionally, we assume that we have resources to take n independent observations in the experiment. The main interest here is to determine an optimal allocation scheme for these n observations in $\mathcal{X}$ to efficiently estimate γ when it is known in advance that some trials might result in non-responses.

Suppose ξ is a design which takes n_j observations at x_j, $j = 1, \ldots, k$, and $\sum_{j=1}^{k} n_j = n$. Following convention, the worth of this design is judged by its Fisher information matrix. If every response is observed, this matrix is proportional to

$$I(\xi, \gamma) = \sum_{j=1}^{k} \frac{n_j}{n} \frac{\partial f(x_j, \gamma)}{\partial \gamma} \frac{\partial f(x_j, \gamma)^T}{\partial \gamma}.$$

If all data are observed, the covariance matrix of the maximum likelihood estimate of γ is proportional to $I(\xi, \gamma)^{-1}$ and optimal designs for estimating γ can be found by maximizing an appropriate function of $I(\xi, \gamma)$. For instance, D-optimal designs for estimating γ are found by maximizing $\det I(\xi, \gamma)$ over the set of designs on $\mathcal{X}$, see for example the design monographs by Fedorov (1972) or Silvey (1980).

In experiments with potentially failing trials, the actual observed sample size may vary from experiment to experiment even for the same design. Our assumption is that at every point x in the design space, there is a known probability $p(x)$ of observing a response in a trial at x. If N_j responses are actually observed out of n_j trials at x_j, $j = 1, \ldots, k$, the observed Fisher information matrix is

$$I^O(\xi, \gamma) = \sum_{j=1}^{k} \frac{N_j}{n} \frac{\partial f(x_j, \gamma)}{\partial \gamma} \frac{\partial f(x_j, \gamma)^T}{\partial \gamma},$$

ignoring an unimportant multiplicative constant. The covariance matrix of the maximum likelihood estimate of γ is proportional to $I^O(\xi, \gamma)^{-1}$ and we are led to finding a design that optimizes $I^O(\xi, \gamma)$ in some sense. However, because N_j is random, $I^O(\xi, \gamma)$ is also random. Consequently, we compare designs using the expected information matrix

$$J(\xi, \gamma) = E\left\{I^O(\xi, \gamma)\right\},$$

where the expectation is taken with respect to the response probabilities at the support points of ξ. If $p(x) = 1$ for all $x \in \mathcal{X}$, so that all observations will be realized, the expected information matrix coincides with the usual information matrix. In practice, we expect that the response probability function $p(x)$ is usually monotonic. For instance, in a chemical experiment, instruments are more likely to malfunction or fail as experimental conditions become more extreme.

Following Herzberg and Andrews (1976) and Hackl (1995), we assume that the individual trials succeed or fail independently of each other. Then, for each j, N_j is a binomial random variable with parameters n_j and $p(x_j)$. Thus $E(N_j) = n_j p(x_j)$ and it follows that

$$J(\xi, \gamma) = \int p(x) \frac{\partial f(x, \gamma)}{\partial \gamma} \frac{\partial f(x, \gamma)^T}{\partial \gamma}\, d\xi(x). \tag{2.2}$$

In what is to follow, we adopt Kiefer's approach (1959) and consider approximate designs only. This means that a design is an arbitrary probability measure on $\mathcal{X}$ with finite support and $n_j = n\xi(x_j)$ is not required to be an integer. We then define the expected information matrix with respect to $p(x)$ by (2.2).

Now suppose that Φ is a user-selected optimality criterion and Φ is a positively homogeneous, concave, increasing and upper semicontinuous function on the set of non-negative matrices, cf. Pukelsheim (1993, Chap.5). The Φ-optimal design for (2.1) with potentially failing trials is the design that maximizes $\Phi(J(\xi,\gamma))$. If the regression model is linear, $J(\xi,\gamma)$ does not depend on γ and we write simply $J(\xi)$. If the model is non-linear, the optimality criterion contains the unknown parameters γ and the designs are locally optimal (Chernoff (1953)). This feature also occurs when there are no failing trials. Locally optimal designs are useful as a first step in designing an experiment for non-linear models, see Ford, Torsney and Wu (1992). In the present situation, locally Φ-optimal designs are found by maximizing $\Phi(J(\xi,\gamma))$ for a nominal value of γ. If there are unknown parameters in the response probability function, one may also specify nominal values for these parameters, and then use a local optimality approach. Alternatively, one may follow a maximin approach to overcome the dependence of the optimal design on the unknown parameters. This is discussed more fully in Section 4.

In the next few sections, we construct optimal designs under various assumptions. Some of these designs are optimal in the sense that the optimal design is found among all designs on $\mathcal{X}$, and others are optimal only among all designs on $\mathcal{X}$ with m points. To distinguish between the two types of optimal designs, the latter designs are called optimal m-point designs. The issue of determining whether these two types of optimal designs are the same is a difficult one; see Dette and Wong (1998). In general, the optimal designs have to be determined using numerical optimization routines. For certain design criteria, standard design algorithms, such as those described in Silvey (1980, Chap.4) can be modified to generate the desired optimal designs. The following sections describe some situations where analytical optimal designs can be found for experiments with potentially failing trials.

3. Optimal Designs for Polynomial Regression

Consider the polynomial model

$$y(x) = \gamma_0 + \gamma_1 x + \cdots + \gamma_m x^m + \epsilon, \qquad x \in \mathcal{X}, \tag{3.1}$$

where $\mathcal{X} \subset \mathbb{R}$ is a compact interval. Our goal is to determine D-optimal designs for this model when there are potentially failing trials and the response probability function $p(x)$ is given. We assume $p(x)$ has the form

$$p(x) = \frac{\kappa}{|x - \theta|}, \tag{3.2}$$

where $\theta \in \mathbb{R} \setminus \mathcal{X}$ and $\kappa > 0$ is so small that $p(x) \le 1$ for all $x \in \mathcal{X}$. This class of response probability functions is flexible because depending on the value of θ, $p(x)$ can be monotonic increasing or decreasing. There is no loss of generality in assuming that $\mathcal{X} = [0, b]$ and θ lies to the left of the regression interval.

Theorem 3.1. *Consider the polynomial regression model* (3.1) *with* $\mathcal{X} = [0, b]$ *and response probability given by* (3.2) *for some* $\theta < 0$. *Then the D-optimal design for experiments with potentially failing trials puts equal masses* $\frac{1}{m+1}$ *at the zeros of the polynomial*

$$x(b - x) \sum_{k=0}^{m-1} c_k x^k, \tag{3.3}$$

where the coefficients are given by

$$c_k = (-1)^k \frac{\prod_{j=0}^{k}(m^2 - j^2)}{k!(k+1)!b^k} \left\{ 2(m^2 - k - 1) + \frac{kb}{\theta} - \frac{k}{\theta}\sqrt{b^2 + 4m^2\theta(\theta - b)} \right\}.$$

Sketch of Proof. Using the theory of oscillatory matrices one may show as in Imhof, Krafft and Schaefer (1998) that the zeros of (3.3) are the sought design points if $(c_0, \ldots, c_{m-1})$ is a characteristic vector of the matrix

$$\begin{bmatrix} r_0 & s_0 & 0 & \cdots & 0 & 0 \\ q_1 & r_1 & s_1 & \cdots & 0 & 0 \\ 0 & q_2 & r_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & q_{m-1} & r_{m-1} \end{bmatrix}, \quad \text{where} \quad \begin{aligned} q_k &= k^2 - m^2, \\ r_k &= -k(kb + k\theta + 3\theta), \\ s_k &= (k+1)(k+2)\theta b. \end{aligned}$$

That this is indeed the case can be verified by a simple but lengthy calculation. The corresponding characteristic value is $\lambda = (2 - m^2)\theta - \frac{b}{2} + \frac{1}{2}\sqrt{b^2 + 4m^2\theta(\theta - b)}$.

It is intuitively reasonable and obvious from Theorem 3.1 that the optimal designs do not depend on κ but do depend on θ. If $\theta < 0$, the response probability is decreasing on $[0, b]$, and if θ is close to zero, then the probability of having a response is much larger at 0 than at b. This suggests that if θ moves towards zero, the design points should move to where the response probability becomes large, that is, to the left. The next theorem confirms this and describes the limit behavior of the design when θ approaches $-\infty$ or $0-$.

Let $G_k(\mu, \nu, x)$ denote the kth monic Jacobi polynomial orthogonal with respect to $(1-x)^{\mu-\nu}x^{\nu-1}$, $x \in [0,1]$; see Abramowitz and Stegun (1965, p.775).

Theorem 3.2. *Every interior support point of the optimal design given in Theorem* 3.1 *moves to the left when $\theta(<0)$ moves to the right. They converge to the zeros of $G_{m-1}(3,2,x/b)$ when $\theta \to -\infty$, and they converge to the zeros of*

$$2G_{m-1}\left(3,2,\frac{x}{b}\right) + (m-1)\frac{x}{b}G_{m-2}\left(4,3,\frac{x}{b}\right)$$

when $\theta \to 0-$.

The proof is given in the appendix. Note that the zeros of $G_{m-1}(3,2,x/b)$ are just the interior support points of the D-optimal design for polynomial regression on $[0,b]$ without failing trials.

4. Maximin Optimal Designs for Partially Specified Response Probabilities

In this section we extend the concepts in Section 2 to situations where the response probability $p(x)$ is not completely known. We limit our discussion to linear models of the form $y(x) = f^T(x)\gamma + \epsilon$, $x \in \mathcal{X}$. We assume only that $p(x)$ belongs to a given class $\{p_\theta(x) : \theta \in \Theta\}$ with a known parameter set Θ. The expected Fisher information matrix $J_\theta(\xi) = \int p_\theta(x)f(x)f^T(x)\,d\xi(x)$ depends now on θ. Let Φ be the underlying optimality criterion. The standardized (cf. Dette (1997)) maximin Φ-optimal design is the design that maximizes

$$\min_{\theta\in\Theta} \frac{\Phi(J_\theta(\xi))}{\max_\eta \Phi(J_\theta(\eta))}. \tag{4.1}$$

When Θ consists of two points, this optimization problem falls within the geometric framework developed by Imhof and Wong (2000), see also Haines (1995).

To illustrate this optimality concept we consider (3.1) with $\mathcal{X} = [-1,1]$ and assume, as Herzberg and Andrews (1976) did, a symmetric response probability function. Specifically, we suppose that the response probabilities have the form

$$p_\theta(x) = (1-x^2)^\theta, \qquad \theta_1 \le \theta \le \theta_2, \tag{4.2}$$

where $\theta_1 < \theta_2$ are fixed known positive numbers. This means that at $x=0$ a response is certain, and near the end-points of $\mathcal{X}$ the response probability decreases to zero. How fast it decreases, however, is only roughly known. We are interested in estimating all the parameters $\gamma_0, \ldots, \gamma_m$ and choose the D-optimality criterion $\Phi(J_\theta(\xi)) = \{\det J_\theta(\xi)\}^{1/(m+1)}$. Then maximizing (4.1) is equivalent to maximizing

$$\min_{\theta_1\le\theta\le\theta_2} \frac{\det J_\theta(\xi)}{\max_\eta \det J_\theta(\eta)},$$

where, according to Karlin and Studden (1966, p.330),

$$\max_{\eta} \det J_\theta(\eta) = 2^{(m+1)(m+2\theta)} \prod_{k=1}^{m} k^k \prod_{k=0}^{m} \frac{(k+\theta)^{2k+2\theta}}{(m+k+2\theta)^{m+k+2\theta}}. \tag{4.3}$$

In the next theorem, we present optimal $(m+1)$-point designs for this setup. The justifications are deferred to the appendix. Investigations in a related context (Dette and Wong (1998), Imhof (2001)) suggest that these designs should be optimal or close to the optimal designs, when the range of θ is not too large.

Theorem 4.1. *The standardized maximin D-optimal $(m+1)$-point design for model* (3.1) *with $\mathcal{X} = [-1,1]$ and response probability structure* (4.2) *puts equal masses at the zeros of the ultraspherical polynomial $P_{m+1}^{(\lambda^*)}(x)$, where $\lambda^* > -1/2$ is uniquely determined by the equation*

$$\left(\prod_{k=0}^{m} \frac{m+k+2\lambda^*+1}{k+\lambda^*+1/2}\right)^{\theta_2-\theta_1} = \prod_{k=0}^{m} \frac{(k+\theta_2)^{k+\theta_2}}{(k+\theta_1)^{k+\theta_1}} \frac{(m+k+2\theta_1)^{\frac{m+k}{2}+\theta_1}}{(m+k+2\theta_2)^{\frac{m+k}{2}+\theta_2}}. \tag{4.4}$$

5. Discussion

When there are potentially failing trials in an experiment, the usual optimality concepts are inappropriate because they provide meaningful comparison only among designs with the same number of valid observations. For a given design criterion Φ, our optimal design maximizes $\Phi(J(\xi,\gamma))$, where J is the expected information matrix with respect to the response probabilities at the support points of the design ξ. In this paper, we have used the D-optimality criterion to illustrate the concepts but the technique can be applied to find other types of optimal designs. For example, E-optimal designs can be similarly derived by combining the ideas of Section 2 with the general theorem on E-optimal designs for Chebyshev systems of Imhof and Studden (2001). The results in Section 4 can also be extended to Bayesian optimality criteria. If we assume that a prior distribution ψ on the parameter θ in the response probability function is available, we may wish to find a Bayesian Φ-optimal design with respect to ψ, which maximizes

$$\int \frac{\Phi(J_\theta(\xi))}{\max_\eta \Phi(J_\theta(\eta))}\, d\psi(\theta).$$

Our experience with several numerical examples, not reported here, suggests that ignoring the possibility of missing observations in the trials can result in a substantial loss of efficiency of the usual optimal design. The gain in efficiency of the D-optimal design proposed here over the usual D-optimal design depends on the model, and generally is not affected by the magnitude of the response probability, but is influenced by the heterogeneity of the non-response structure.

We conclude with a note that in experiments with possibly failing trials, it is possible that the observed information matrix $I^O(\xi,\gamma)$ is singular, even if ξ is so chosen that the ordinary information matrix $I(\xi,\gamma)$ is non-singular. In this case, not all the parameters in γ are estimable and it is interesting to compare the probability of this occurrence using the proposed design and the usual optimal design. To do this, we first fix a nominal value of γ and assume that $n\xi(x_j)$ is again an integer for each design point x_j, $j = 1,\ldots,k$. We also assume that the entries of the vector $\partial f(x,\gamma)/\partial\gamma$ form a Chebyshev system on $\mathcal{X}$, see Karlin and Studden (1966). Then $I^O(\xi,\gamma)$ is singular if and only if the number of design points with at least one valid response is less than m, the dimension of γ. Because the $N_i's$ are independent binomial random variables with parameters $n\xi(x_i)$ and $p(x_i)$, it now follows that

$$P\left\{I^O(\xi,\gamma) \text{ is singular}\right\} = \sum_{j=0}^{m-1} \sum_{\substack{S\subset\{1,\ldots,k\}\\ |S|=j}} P\left(N_i > 0 \text{ if } i \in S; N_i = 0 \text{ if } i \notin S\right)$$
$$= \sum_{j=0}^{m-1} \sum_{\substack{S\subset\{1,\ldots,k\}\\ |S|=j}} \prod_{i\in S}\left[1-\{1-p(x_i)\}^{n\xi(x_i)}\right] \prod_{i\notin S}\{1-p(x_i)\}^{n\xi(x_i)}. \tag{5.1}$$

Note that the last expression makes sense even if $n\xi(x_i)$ is not an integer. Given any design ξ, we use (5.1) to calculate the minimal value of n which ensures that the probability of observing a singular information matrix is below a prescribed level. If the entries of $\partial f(x,\gamma)/\partial\gamma$ do not form a Chebyshev system, the right-hand side of (5.1) still gives a lower bound for the probability that $I^O(\xi,\gamma)$ is singular, because the matrix may, then, be singular even if responses at m or more different points in $\mathcal{X}$ have been obtained.

The optimal designs proposed here tend to have smaller probabilities that their observed information matrices are singular when compared with those from the usual optimal designs. Table 1 gives an example of the results obtained when we compare the probabilities that the usual optimal design and our proposed optimal design each has a non-singular information matrix. The probabilities are computed using the D-optimal designs found in Section 3 for the cubic model on $[0, 1]$ assuming different nominal values in the response probability function. The sample size is 80 and the usual D-optimal design ξ^* for the cubic model is supported at $0.0, 0.276, 0.724$ and 1.0. The table shows that our optimal designs consistently have a higher probability of producing a non-singular observed information matrix than the usual optimal design. That this is in fact always the case for the present model follows from Theorem 3.2. This suggests that our designs may have an additional advantage over the usual optimal designs in experiments with potentially failing trials. Further research in this direction is underway.

Table 1. Support of the D-optimal design ξ for cubic regression on $[0,1]$ with $p(x) = \kappa/|x-\theta|$. The last two columns list the probabilities of obtaining a non-singular observed information matrix using ξ and the usual optimal design ξ^*, respectively.

θ	κ	supp(ξ)	$P(I^O(\xi)$ non-s.)	$P(I^O(\xi^*)$ non-s.)
-0.1	0.10	$\{0.0, 0.197, 0.665, 1.0\}$	0.799	0.786
-0.5	0.25	$\{0.0, 0.238, 0.691, 1.0\}$	0.965	0.963
-1.0	0.30	$\{0.0, 0.252, 0.702, 1.0\}$	0.937	0.935

Acknowledgements

We thank the associate editor and referees for helpful comments, and Prof. Peter Hackl for additional references in this work. The research of Wong is partially supported by an NIH research grant R29 AR44177-01A1.

Appendix

Proof of Theorem 3.2. Let $c_k = c_k(\theta)$ be as in Theorem 3.1. Then

$$\sum_{k=0}^{m-1} c_k(\theta)x^k = (-1)^{m+1}\frac{(2m-1)!}{(m-1)!(m-2)!}h(x,\theta),$$

$$h(x,\theta) = 4mG_{m-1}\Big(3,2,\frac{x}{b}\Big)+\frac{x}{b}\Big\{-2m+\frac{b}{\theta}-\frac{\sqrt{b^2+4m^2\theta(\theta-b)}}{\theta}\Big\}G_{m-2}\Big(4,3,\frac{x}{b}\Big).$$

The term in braces converges to 0 as $\theta \to -\infty$ and to $-2m+2m^2$ as $\theta \to 0-$. This proves the limit assertions.

Now let $x_1(\theta) < \cdots < x_{m-1}(\theta)$ be the interior design points, i.e., the zeros of $h(x,\theta)$. By the Implicit Function Theorem, each $x_k(\theta)$ is differentiable and

$$\frac{dx_k(\theta)}{d\theta} = -\frac{h_2(x_k(\theta),\theta)}{h_1(x_k(\theta),\theta)} = \frac{x_k(\theta)}{\theta^2}\Big\{1-\frac{b-2m^2\theta}{\sqrt{b^2+4m^2\theta(\theta-b)}}\Big\}\frac{G_{m-2}(4,3,x_k(\theta)/b)}{h_1(x_k(\theta),\theta)}.$$

The term in braces is seen to be negative and $\text{sgn}h_1(x_k(\theta),\theta) = (-1)^{m+k-1}$. To determine $\text{sgn}G_{m-2}(4,3,x_k(\theta)/b)$, let $\xi_1 < \cdots < \xi_{m-2}$ and $\eta_1 < \cdots < \eta_{m-1}$ denote the zeros of $G_{m-2}(4,3,x/b)$ and $G_{m-1}(3,2,x/b)$, respectively. Let $\xi_0 = 0$, $\xi_{m-1} = b$. An application of Sturm's comparison theorem (Szegö (1975, p.19)) to the differential equation (4.24.2) in Szegö (1975, p.67) shows that $\xi_0 < \eta_1 < \xi_1 <$

$\cdots < \eta_{m-1} < \xi_{m-1}$. Therefore, $\mathrm{sgn} h(\xi_k, \theta) = \mathrm{sgn} G_{m-1}(3, 2, \xi_k/b) = (-1)^{m+k-1}$ for $k = 0, \ldots, m-1$. Thus $x_k(\theta) \in (\xi_{k-1}, \xi_k)$, so that $\mathrm{sgn} G_{m-2}(4, 3, x_k(\theta)/b) = (-1)^{m+k-1}$. Hence $dx_k(\theta)/(d\theta) < 0$.

Proof of Theorem 4.1. A standard design argument (see, e.g., Silvey (1980, p.43) shows that we can restrict attention to designs which put equal masses on their $m + 1$ support points. Thus let ξ be a design with support points $x_0 < \cdots < x_m$ in $(-1, 1)$ and let $\xi(x_i) = 1/(m + 1)$ for $i = 0, \ldots, m$. Then

$$\det J_\theta(\xi) = \frac{1}{(m+1)^{m+1}} \prod_{k=0}^{m} (1 - x_k^2)^\theta \prod_{i<j} (x_j - x_i)^2$$

for all $\theta \in \Theta$. Now define the function

$$H(\lambda) = \left(\prod_{k=1}^{m+1} \frac{2k + 2\lambda - 1}{m + k + 2\lambda} \right)^2, \qquad -\frac{1}{2} \leq \lambda.$$

This function is strictly increasing on $[-\frac{1}{2}, \infty)$ with $H(-\frac{1}{2}) = 0$ and $\lim_{\lambda\to\infty} H(\lambda) = 1$. There exists, therefore, a unique $\lambda_\xi > -\frac{1}{2}$ such that $H(\lambda_\xi) = \prod_{k=0}^{m}(1 - x_k^2)$. Using Theorem 3.2 in Karlin and Studden (1966, p.330) and formulas (4.7.3) and (4.7.9) in Szegö (1975, p.80f) one may show that

$$\det J_\theta(\xi) \leq 2^{m(m+1)} \{H(\lambda_\xi)\}^\theta \prod_{k=1}^{m} k^k \prod_{k=1}^{m+1} \frac{(k + \lambda - 1/2)^{2k-2}}{(m + k + 2\lambda)^{m+k-1}}.$$

Thus, by (4.3),

$$\frac{\det J_\theta(\xi)}{\max_\eta \det J_\theta(\eta)} \leq \prod_{k=0}^{m} \left(\frac{k+\lambda_\xi+1/2}{k+\theta} \right)^{2k+2\theta} \left(\frac{m + k + 2\theta}{m + k + 2\lambda_\xi+1} \right)^{m+k+2\theta}, \qquad \text{(A.1)}$$

and there is equality if and only if $x_0, \ldots, x_m$ are the zeros $P_{m+1}^{(\lambda_\xi)}(x)$. Let $K(\theta, \lambda)$ denote the expression on the right side of (A.1) with λ_ξ replaced by λ. As a function of θ, $K(\theta, \lambda)$ is strictly increasing for $0 < \theta \leq \lambda + \frac{1}{2}$ and strictly decreasing for $\theta \geq \lambda + \frac{1}{2}$. Thus

$$\min_{\theta_1 \leq \theta \leq \theta_2} \frac{\det J_\theta(\xi)}{\max_\eta \det J_\theta(\eta)} \leq \min_{\theta_1 \leq \theta \leq \theta_2} K(\theta, \lambda_\xi) = \min\{K(\theta_1, \lambda_\xi), K(\theta_2, \lambda_\xi)\}.$$

Since $K\left(\theta_1, \theta_1 - \frac{1}{2}\right) = 1 > K\left(\theta_2, \theta_1 - \frac{1}{2}\right)$ and $K\left(\theta_2, \theta_2 - \frac{1}{2}\right) = 1 > K\left(\theta_1, \theta_2 - \frac{1}{2}\right)$, there is some $\lambda \in (\theta_1 - \frac{1}{2}, \theta_2 - \frac{1}{2})$ such that $K(\theta_1, \lambda) = K(\theta_2, \lambda) =: K^*$, say. This λ is just λ^*, which in particular ensures that (4.4) has indeed a solution. Moreover, $\min\{K(\theta_1, \lambda), K(\theta_2, \lambda)\} < K^*$ if $\lambda \neq \lambda^*$. It therefore follows that ξ is the optimal design if and only if $x_0, \ldots, x_m$ are the zeros of $P_{m+1}^{(\lambda^*)}(x)$.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions.* Dover, New York.

Akhtar, M. and Prescott, P. (1986). Response surface designs robust to missing observations. *Comm. Statist. Simulation Comput.* **15**, 345-363.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.

Dette, H. (1997). Designing experiments with respect to 'standardized' optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59**, 97-110.

Dette, H. and Wong, W. K. (1998). Bayesian D-optimal designs on a fixed number of design points for heteroscedastic polynomial models. *Biometrika* **85**, 869-882.

Fedorov, V. V. (1972). *Theory of Optimal Experiments.* Translated and edited by W. J. Studden and E. M. Klimko. Academic Press, New York.

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.

Hackl, P. (1995). Optimal design for experiments with potentially failing trials. In *Proc. of MODA4: Advances in Model-Oriented Data Analysis* (Edited by C. P. Kitsos and W. G. Müller), 117-124. Physica Verlag, Heidelberg.

Haines, L. M. (1995). A geometric approach to optimal design for one-parameter non-linear models. *J. Roy. Statist. Soc. Ser. B* **57**, 575-598.

Herzberg, A. M. and Andrews, D. F. (1976). Some considerations in the optimal design of experiments in non-optimal situations. *J. Roy. Statist. Soc. Ser. B* **38**, 284-289.

Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Ann. Statist.* **29**, 561-576.

Imhof, L., Krafft, O. and Schaefer, M. (1998). D-optimal designs for polynomial regression with weight function $x/(1+x)$. *Statist. Sinica* **8**, 1271-1274.

Imhof, L. A. and Studden, W. J. (2001). E-optimal designs for rational models. *Ann. Statist.* **29**, 763-783.

Imhof, L. and Wong, W. K. (2000). A graphical method for finding maximin efficiency designs. *Biometrics* **56**, 113-117.

Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Interscience, New York.

Kiefer, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **21**, 272-319.

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

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

Szegö, G. (1975). *Orthogonal Polynomials.* 4th ed. Amer. Math. Soc. Colloq. Publ. 23. Amer. Math. Soc., Providence, RI.

Institut für Statistik, RWTH Aachen, D-52056 Aachen, Germany.

E-mail: imhof@stochastik.rwth-aachen.de

Scirex Corporation, 255 E Lake Street, Bloomingdale, Illinois 60108, U.S.A.

E-mail: dsong@scirex.com

Department of Biostatistics, University of California, Los Angeles, California 90095-1772, U.S.A.

E-mail: wkwong@sunlab.ph.ucla.edu

(Received March 2001; accepted May 2002)