OPTIMAL DESIGNS FOR RATIONAL MODELS

By Zhuoqiong He,[1] William J. Studden[2] and Dongchu Sun[3]

*Missouri Department of Conservation, Purdue University
and University of Missouri–Columbia*

In this paper, experimental designs for a rational model, $Y = P(x)/Q(x) + \varepsilon$, are investigated, where $P(x) = \theta_0 + \theta_1 x + \cdots + \theta_p x^p$ and $Q(x) = 1 + \theta_{p+1} x + \cdots + \theta_{p+q} x^q$ are polynomials and ε is a random error. Two approaches, Bayesian D-optimal and Bayesian optimal design for extrapolation, are examined. The first criterion maximizes the expected increase in Shannon information provided by the experiment asymptotically, and the second minimizes the asymptotic variance of the maximum likelihood estimator (MLE) of the mean response at an extrapolation point x_e. Corresponding locally optimal designs are also discussed. Conditions are derived under which a $p + q + 1$-point design is a locally D-optimal design. The Bayesian D-optimal design is shown to be independent of the parameters in $P(x)$ and to be equally weighted at each support point if the number of support points is the same as the number of parameters in the model. The existence and uniqueness of the locally optimal design for extrapolation are proven. The number of support points for the locally optimal design for extrapolation is exactly $p + q + 1$. These $p + q + 1$ design points are proved to be independent of the extrapolation point x_e and the parameters in $P(x)$. The corresponding weights are also independent of the parameters in $P(x)$, but depend on x_e and are not equal.

1. Introduction. Experimental design has its root in the agricultural experiments of Fisher's pioneering work at Rothamsted Experimental Station in the 1920s and 1930s, and has flourished as an integral part of statistical research and practice. Most of the design work has focused on the linear regression model due to its simplicity. However, as the amount of statistical analysis using nonlinear models increased in fields such as the chemical, biological and clinical sciences, people became aware of the need for optimal design for these nonlinear models.

The earliest relevant nonlinear example is Fisher's (1922) work on the dilution-series problem; see Fedorov (1972), page 121. The next major impetus took the form of various nonlinear regression models involving exponential functions for reaction rates in chemical kinetics; examples are included in Box and Lucas (1959) and Katz, Azen and Schumitzky (1981). Another important class of nonlinear regression models is the rational model, which has been

widely used in agricultural and biological research; see, for example, Sparrow (1979a, b).

The nonlinear problem is intrinsically much harder than the linear problem as the information matrix depends on the unknown parameters. A common approach is to design an experiment to be efficient for a best guess of the parameter values. This approach leads to what are called "locally optimal" designs, introduced by Chernoff (1953) and discussed in Fedorov [(1972), Sections 1.4, 2.8 and 4.4], Silvey [(1980), Sections 6 and 7], and Ford, Titterington and Kitsos (1989). Although, in practice, parameters are rarely known, Ford, Torsney and Wu (1992) note various reasons why locally optimal designs are still of interest.

A natural generalization of the locally optimal design is to use a prior distribution on the parameters rather than a single guess. An optimal design is then called a Bayesian optimal design. Although numerous articles have been written on the classical aspects of nonlinear models, such as estimation and inference, the area of nonlinear experimental designs, especially Bayesian nonlinear experimental design, started receiving attention much more recently. The literature in this area includes Tsutakawa (1972), Chaloner and Larntz (1989) and Chaloner (1993). Finding the Bayesian D-optimal design for a nonlinear model is usually quite difficult. Chaloner (1993) gives a closed form of the Bayesian D-optimal design for a one-parameter logistic model with a prior having two equally weighted points. For nondegenerate prior distributions, the examples of Bayesian D-optimal designs considered so far for models with more than one parameter are found numerically and, in general, cannot be expressed in closed form.

It is well known that various versions of the Kiefer–Wolfowitz equivalence theorem play an extremely important role in almost all the nonlinear design problems, including both theoretical and numerical approaches. Such equivalence theorems for nonlinear optimal designs can be found in Whittle (1973) and Läuter (1974), and for Bayesian nonlinear designs in Chaloner and Larntz (1989) and Pilz (1991). One usually resorts to finding an optimal design within the class of k-point designs for a particular value of k and then uses the equivalence theorem to check that the resulting design is indeed globally optimal.

The article is organized as follows. In Section 2, we review several Bayesian optimal design criteria for the nonlinear problem, and discuss some basic properties about these designs. We will illustrate that the Bayesian D-optimal design under any prior distribution should be equally weighted at each support point if the number of support points is the same as the number of parameters in the model.

In Section 3, we consider a general rational model

$$Y = \frac{\theta_0 + \theta_1 x + \cdots + \theta_p x^p}{1 + \theta_{p+1} x + \cdots + \theta_{p+q} x^q} + \varepsilon, \qquad x \in [a, b].$$

We will prove that, under some suitable conditions, the Bayesian D-optimal and optimal designs for extrapolation depend only on the marginal distributions of $(\theta_{p+1}, \ldots, \theta_{p+q})$.

In Section 4, we show that locally D-optimal designs for rational models are equally weighted on exactly $k = p + q + 1$ support points under some mild conditions about the parameters. The relationships between rational models and polynomial models are also sought.

In Section 5, we construct the locally optimal design for extrapolation, which is supported on $p + q + 1$ points and is independent of both the parameters $(\theta_0, \theta_1, \ldots, \theta_p)$ and the extrapolation point x_e. If $p \geq q$, these support points are $(a = s_0 < s_1 < \cdots < s_{p+q-1} < s_{p+q} = b)$, where $(s_1, \ldots, s_{p+q-1})$ are determined by the 0's of a certain polynomial. This is analogous to the result of Hoel and Levine (1964) for the ordinary polynomial case.

Finally, the Michaelis–Menten model is considered in Section 6. The analytic form of the D-Bayesian optimal two-point designs is discussed. The locally D-optimal and optimal designs for extrapolation are given and compared.

2. Optimal designs for a general nonlinear model.

2.1. *A nonlinear regression model.* To introduce the notation, consider a general nonlinear regression model of the form:

$$(1) \qquad\qquad Y = Y(x) = f(x, \boldsymbol{\theta}) + \varepsilon,$$

where Y is the response variable, x is an explanatory variable, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ is a k vector of unknown parameters, $f = f(x, \boldsymbol{\theta})$, the mean response at x, is a nonlinear function of $\boldsymbol{\theta}$ and ε is a random error satisfying $\mathbb{E}(\varepsilon) = 0$ and $var(\varepsilon) = \sigma^2$. Let $\mathscr{X}$ denote the *design space* of possible x's, which is usually a compact subset of $\mathbb{R}^p$. Assume that ε has a normal distribution with mean 0 and variance σ^2. In general, σ^2 may depend on both x and $\boldsymbol{\theta}$. However, for simplicity, we assume that σ^2 is a known constant and independent of x and $\boldsymbol{\theta}$. This is usually not too restrictive since our main purpose is to design the experiment concerning the parameter $\boldsymbol{\theta}$. Without loss of generality, we assume that $\sigma^2 = 1$.

Associated with an observation, Y, there is a Fisher information matrix, $I(\boldsymbol{\theta}, x)$, given by

$$(2) \qquad
\begin{aligned}
I(\boldsymbol{\theta}, x) &= \nabla f (\nabla f)^t, \\
\nabla f \equiv \nabla f(x, \boldsymbol{\theta}) &= \left(\frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \ldots, \frac{\partial f}{\partial \theta_k} \right)^t.
\end{aligned}$$

Any design can be represented by a probability measure ξ on $\mathscr{X}$. If ξ is supported at m points $x_1, \ldots, x_m$ with mass $\xi(x_i) = n_i/n$, $\sum_{i=1}^{m} n_i = n$, the experiment takes n_i observations at x_i. The average (per observation) information matrix of a design ξ is then defined by

$$(3) \qquad\qquad M(\xi) \equiv M(\boldsymbol{\theta}; \xi) = \int_{\mathscr{X}} I(\boldsymbol{\theta}, x) \xi(dx).$$

Let ξ_n denote the n-point design taking observations at the points $\mathbf{x}_{(n)} = (x_1, \ldots, x_n)$. If ε is not normally distributed but all the observations are mutu-

ally independent, then $M^{-1}(\theta; \xi_n)$ is proportional to the asymptotic covariance matrix for the maximum likelihood estimate for θ, provided that $M^{-1}(\theta; \xi_n)$ exists.

2.2. *The design criteria.* Bayesian optimal designs minimize the expectation of some function of $M(\theta; \xi)$ with respect to some prior distribution. The most commonly used criterion is

$$(4) \qquad \phi^D(\xi) = \int_\Theta \log\{\det[M(\theta; \xi)]\}\, d\pi(\theta),$$

where $\pi(\cdot)$ is a prior distribution of θ. This criterion was originally suggested by Lindley (1956) and used recently by Chaloner and Larntz (1989) and Chaloner (1993), for instance. Under an asymptotic normal approximation to the posterior distribution of θ, $\phi^D(\cdot)$ approximates the expected increase in Shannon information provided by the experiment. A design ξ minimizing (4) is called a Bayesian D-optimal design. If $\pi(\cdot)$ is degenerate, a Bayesian D-optimal design is referred to as a locally D-optimal design.

Another commonly used design is a Bayesian L-optimal design minimizing

$$(5) \qquad \phi^L(\xi) = \int_\Theta \mathrm{tr}\{L(\theta)M^{-1}(\theta; \xi)\}\, d\pi(\theta),$$

where $L(\theta)$ is a symmetric, nonnegative definite, $k \times k$ matrix. If linear combinations of the θ_i's are of interest, then $L(\theta)$ does not depend on θ and is a constant matrix. If nonlinear combinations of the θ_i's are of interest, then $L(\theta)$ has entries which are functions of θ. If $\pi(\cdot)$ is a degenerate prior, the Bayesian L-optimal design is referred to as the locally L-optimal design.

When L is a rank-one matrix of the form $L = \mathbf{cc}'$ for some $k \times 1$ column vector $\mathbf{c}$, the corresponding designs are referred to as c-optimal.

In design problems, the experimenter often wants to know the dependence of the response on explanatory variables in those regions where observations, from the practical point of view, are impossible or extremely difficult to collect. To be specific, suppose that we need to extrapolate to a given point $x_e \notin \mathscr{X}$. This means estimating $f(x_e, \theta)$, the mean response at x_e. Let $\hat{\theta}$ be the MLE of θ. For any design ξ, the variance of the asymptotic distribution of $f(x_e, \hat{\theta})$ for fixed θ is then $\nabla f(x_e, \theta)^t M^{-1}(\theta, \xi) \nabla f(x_e, \theta)$. This motivates a special c-optimal design criterion with $\mathbf{c} = \nabla f(x_e, \theta)$. A design ξ minimizing such a criterion will be called a Bayesian c_e-optimal design. The corresponding locally optimal design will be called the c_e-locally optimal design.

The following theorem gives an upper bound for the number of support points needed for a Bayesian optimal design.

THEOREM 1. *If the prior distribution of θ is supported on no more than r distinct points, there exists a Bayesian D-optimal design, supported at no more than $\frac{1}{2}rk(k+1)$ points.*

PROOF. This can be proven in a manner similar to Theorem 4 in Läuter (1974). See also Theorem 2.1 of Dette and Neugebauer (1995). □

In the case of one-parameter nonlinear models, this means that the support of the Bayesian optimal designs consists of no more points than the support of the prior distribution. In addition, there must be a locally optimal design with one support point.

THEOREM 2.   *Consider a general nonlinear regression model, $Y = f(x, \boldsymbol{\theta}) + \varepsilon$, where $x \in \mathscr{X} = [a, b]$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. For any fixed prior of $\boldsymbol{\theta}$, among all the k-point designs, a Bayesian D-optimal design has equal weight on its k supporting points.*

PROOF.   See Karlin and Studden (1966), page 329. □

We should point out that for a k-parameter design problem, if the Bayesian D-optimal design ξ has more than k supporting points, ξ may not be equally weighted.

3. Bayesian optimal designs.

3.1. *Preliminaries and a gradient identity.*   In this section, we consider a general rational regression model,

(6) $Y = f(x, \boldsymbol{\theta}) + \varepsilon$,   with $f(x, \boldsymbol{\theta}) = P(x)/Q(x)$ and $x \in \mathscr{X} \equiv [a, b]$,

where $P(x) = P(x, \boldsymbol{\theta}_P)$ and $Q(x) = Q(x, \boldsymbol{\theta}_Q)$ are two polynomials in x, defined by

(7) $P(x) = \theta_0 + \theta_1 x + \cdots + \theta_p x^p,$

(8) $Q(x) = 1 + \theta_{p+1} x + \cdots + \theta_{p+q} x^q.$

Here $k = p + q + 1$, $\boldsymbol{\theta}_P = (\theta_0, \theta_1, \ldots, \theta_p)^t$, $\boldsymbol{\theta}_Q = (\theta_{p+1}, \ldots, \theta_{p+q})^t$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_P, \boldsymbol{\theta}_Q)$. This, of course, is a special case of the general nonlinear regression model (1). We will examine the design problems for the Bayesian D-optimal design and the Bayesian c_e-optimal design. More detailed results about the corresponding locally optimal designs will be investigated in Sections 4 and 5.

In using Kiefer–Wolfowitz-type equivalence theorems, a useful concept is the directional derivative of ξ in the direction ξ_x, defined by

$$d(\xi, x) = \lim_{\varepsilon \downarrow 0} \frac{\phi((1 - \varepsilon)\xi + \varepsilon \xi_x) - \phi(\xi)}{\varepsilon}.$$

Here ξ_x is a point mass at $x \in [a, b]$ and $\phi(\xi)$ is a design criterion such that $d(\xi, x)$ exists. From Whittle (1973) or Chaloner and Larntz (1989), a design ξ_0 is a Bayesian optimal design if and only if

$$d(\xi_0, x) \leq 0 \quad \text{for all } x \in \mathscr{X}.$$

To obtain an expression for $d(\xi, x)$, we first derive a simple expression for the gradient vector ∇f. For any positive integer n, let

(9) $\mathbf{h}_n \equiv \mathbf{h}_n(x) = (1, x, x^2, \ldots, x^n)^t.$

For any integer i, $0 \le i \le p + q$, let $\mathbf{b}_i$ denote the following $p + q + 1$ column vector:

$$
(10) \quad \mathbf{b}_i =
\begin{cases}
(\underbrace{0, 0, \ldots, 0}_{i}, 1, \theta_{p+1}, \theta_{p+2}, \ldots, \theta_{p+q}, \underbrace{0, 0, \ldots, 0}_{p-i})^t, \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } 0 \le i \le p, \\
-(\underbrace{0, 0, \ldots, 0}_{i-p}, \theta_0, \theta_1, \ldots, \theta_p, \underbrace{0, 0, \ldots, 0}_{p+q-i})^t, \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } p + 1 \le i \le p + q.
\end{cases}
$$

Define a $(p + q + 1) \times (p + q + 1)$ matrix $\mathbf{B}$ by

$$
(11) \qquad\qquad \mathbf{B} \equiv \mathbf{B}(\boldsymbol{\theta}) = (\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_{p+q})^t.
$$

Also let

$$
(12) \qquad g_i(x) = \frac{\partial}{\partial \theta_i} f(x, \boldsymbol{\theta}) =
\begin{cases}
\dfrac{x^i}{Q(x)}, & \text{if } 0 \le i \le p, \\[2mm]
-\dfrac{x^{i-p} P(x)}{Q^2(x)}, & \text{if } p + 1 \le i \le p + q.
\end{cases}
$$

LEMMA 1 (A gradient identity). *We have the following relationship*:

$$
(13) \qquad\qquad \nabla f = \nabla f(x, \boldsymbol{\theta}) = B(\boldsymbol{\theta}) \mathbf{h}_{p+q}^t(x) / Q^2(x, \boldsymbol{\theta}_Q).
$$

The proof of the identity is fairly straightforward and will be omitted. It is crucial in proving most of the results related to the rational model (6).

In the following it will always be assumed that on the support sets of ξ and π the polynomial $Q^2(x, \boldsymbol{\theta}_Q)$ is bounded away from 0 and that $\int \log(|B(\boldsymbol{\theta})|^2) \, d\pi(\boldsymbol{\theta}) < \infty$.

LEMMA 2. *For the rational model* (6), *the Fisher information matrix $I(\boldsymbol{\theta}, x)$ and the average information $M(\boldsymbol{\theta}, \xi)$ are*

$$
(14) \qquad\qquad I(\boldsymbol{\theta}, x) = \frac{1}{Q^4} \mathbf{B} \mathbf{h}_{p+q} \mathbf{h}_{p+q}^t \mathbf{B}^t
$$

and

$$
(15) \qquad\qquad M(\boldsymbol{\theta}; \xi) = \mathbf{B} M_0(\boldsymbol{\theta}_Q; \xi) \mathbf{B}^t,
$$

respectively, where

$$
(16) \qquad M_0(\boldsymbol{\theta}_Q; \xi) = \int_a^b \frac{1}{Q^4(x; \boldsymbol{\theta}_Q)} \mathbf{h}_{p+q}(x) \mathbf{h}_{p+q}^t(x) \xi(dx).
$$

PROOF. This follows immediately from the gradient identity in Lemma 1. $\square$

3.2. *Bayesian D-optimal designs.*

THEOREM 3. *The design criterion $\phi^D(\xi)$ and the corresponding directional derivative $d^D(\xi, x)$ are*

$$(17) \qquad \phi^D(\xi) = \int_\Theta \log(|\mathbf{B}|^2)\pi(d\boldsymbol{\theta}) + \int_\Theta \log(|M_0(\boldsymbol{\theta}_Q; \xi)|)\pi(d\boldsymbol{\theta})$$

and

$$\begin{aligned}
& d^D(\xi, x) \\
(18) \quad & = \int_\Theta \mathrm{tr}\left\{ M_0^{-1}(\boldsymbol{\theta}_Q; \xi)\left(\frac{1}{Q^4}\mathbf{h}_{p+q}(x)\mathbf{h}_{p+q}^t(x) \right) \right\}\pi(d\boldsymbol{\theta}) - (p+q+1),
\end{aligned}$$

respectively.

PROOF. The result follows from the definition of directional derivative and formula (1.1.34) of Fedorov (1972). □

COROLLARY 1. *Under the assumptions of Theorem 2, the Bayesian D-optimal design is independent of $\boldsymbol{\theta}_P = (\theta_0, \theta_1, \ldots, \theta_P)$, the parameters in the numerator polynomial $P(x, \boldsymbol{\theta}_P)$.*

Since the mean response is linear in $\boldsymbol{\theta}_P$, the parameters in $P(x)$, one expects that the optimal design should be independent of $\boldsymbol{\theta}_P$. This type of nonlinear regression model is known as a partially nonlinear model and has been discussed by Hill (1980). It is expected that an optimal design for a partially nonlinear model is independent of the linear parameters. Unfortunately, this is not always the case and a counterexample can be found in Hill (1980). When all the parameters in a rational regression model are of interest, Corollary 1 confirms the above property for the design criterion (4). Surprisingly, if we are interested in only a subset of all the parameters in the model, a Bayesian or locally D-optimal design may depend on all the parameters in the model in general. For details, see He (1993).

REMARK 1. If the Bayesian D-optimal design ξ has support on $p + q + 1$ points, then Theorem 2 tells us that ξ must have equal weight on these points.

EXAMPLE 1. In describing crop yield on the fertilizer input, Sparrow (1979a) considered the model

$$(19) \qquad Y = (\theta_0 + \theta_1 x)/(1 + \theta_2 x) + \varepsilon, \qquad x \in [a, b],$$

where $\theta_2 \geq 0$, $0 \leq a < b < \infty$. This a special case of (6) when $p = q = 1$. It is easy to see that

$$\mathbf{B} = \begin{pmatrix} 1 & \theta_2 & 0 \\ 0 & 1 & \theta_2 \\ 0 & -\theta_0 & -\theta_1 \end{pmatrix}$$

and $|\mathbf{B}| = \theta_0\theta_2 - \theta_1$.

PROPOSITION 1.   *Consider the model* (19). *Assume that the marginal prior distribution of* θ_2 *has half its mass at each of the points* c *and* d, *where* $0 < c \le d$, *and that, on the support of the prior,* $\theta_0 \theta_2 - \theta_1 \neq 0$. *Then, among all three-point designs, the Bayesian D-optimal design* ξ *has equal weight on* $\{a, x^*, b\}$, *where*

(20) $$x^* = \frac{abcd - 1 + \sqrt{(ac+1)(ad+1)(bc+1)(bd+1)}}{(a+b)cd + c + d}.$$

PROOF.   For any design ξ having equal weight at points $a \le x_0 < x_1 < x_2 \le b$,

$$\phi^D(\xi) = C + 2g(x_0, x_1, x_2),$$

where C is some constant and

$$g(x_0, x_1, x_2) = \log \frac{(x_1 - x_0)(x_2 - x_0)(x_2 - x_1)}{(cx_0+1)(cx_1+1)(cx_2+1)(dx_0+1)(dx_1+1)(dx_2+1)}.$$

Using this, it can readily be verified that x_0 must be a and then $x_2 = b$. The derivative with respect to x_1 results in a quadratic equation of which (20) is the relevant root.   □

We have found the Bayesian D-optimal design for the model (19) among all three-point designs. An open question is to verify that this design is indeed the Bayesian D-optimal design.

3.3. *Bayesian L-optimal and* c_e-*optimal designs.*   We first derive the formula for the general Bayesian L-optimal design.

THEOREM 4.   *The design criterion* $\phi^L(\xi)$ *and the corresponding directional derivative* $d^L(\xi, x)$ *are*

(21) $$\phi^L(\xi) = \int_\Theta \text{tr}\{\mathbf{B}^{-1} L(\boldsymbol{\theta}) (\mathbf{B}^{-1})^t M_0^{-1}(\boldsymbol{\theta}_Q; \xi)\} \pi(d\boldsymbol{\theta})$$

and

(22)
$$d^L(\xi, x) = \phi^L(\xi) - \int_\Theta \text{tr}\left\{\mathbf{B}^{-1} L(\boldsymbol{\theta})(\mathbf{B}^{-1})^t M_0^{-1}(\boldsymbol{\theta}_Q; \xi) \right.$$
$$\left. \times \left(\frac{1}{Q^4} \mathbf{h}_{p+q} \mathbf{h}_{p+q}^t\right) M_0^{-1}(\boldsymbol{\theta}_Q; \xi)\right\} \pi(d\boldsymbol{\theta}),$$

respectively. Here $M_0(\boldsymbol{\theta}_Q; \xi)$ *is given by* (16) *and* $\mathbf{B}$ *is given by* (11).

PROOF.   This follows from Lemma 1 above and formula (1.1.33) of Fedorov (1972).   □

Suppose that it is necessary to estimate $f(x_e, \boldsymbol{\theta}) = P(x_e)/Q(x_e)$, the mean response at a given point $x_e \notin [a, b]$. A c_e-optimal design is an L-optimal

design with $L(\boldsymbol{\theta}) = \nabla f(x_e, \boldsymbol{\theta}) \nabla f(x_e, \boldsymbol{\theta})^t = Q^{-4}(x_e) \mathbf{B} \mathbf{h}_{p+q}(x_e) \mathbf{h}_{p+q}^t(x_e) \mathbf{B}^t$. Then we have the following results.

COROLLARY 2. *The Bayesian c_e-optimal design criterion and the corresponding directional derivative are*

$$(23) \quad \phi_e(\xi) = \phi_e(\xi, x_e) = \int_\Theta \frac{1}{Q^4(x_e)} \mathbf{h}_{p+q}^t(x_e) M_0^{-1}(\boldsymbol{\theta}_Q; \xi) \mathbf{h}_{p+q}(x_e) \pi(d\boldsymbol{\theta})$$

and

$$(24) \quad \begin{aligned} d_e(\xi, x) &= \phi_e(\xi) - \int_\Theta \frac{1}{Q^4(x_e) Q^4(x)} \\ &\quad \times \{\mathbf{h}_{p+q}^t(x_e) M_0^{-1}(\boldsymbol{\theta}_Q; \xi) \mathbf{h}_{p+q}(x)\}^2 \pi(d\boldsymbol{\theta}), \end{aligned}$$

respectively.

COROLLARY 3. *If $\mathbf{B}^{-1} L(\boldsymbol{\theta})(\mathbf{B}^{-1})^t$ does not depend on $\boldsymbol{\theta}_P$, then the Bayesian design criterion $\phi^L(\xi)$ does not depend on $\boldsymbol{\theta}_P$. In particular, the Bayesian c_e-optimal design does not depend on the marginal distribution of $\boldsymbol{\theta}_Q$.*

Although the Bayesian D-optimal design itself does not depend on $\boldsymbol{\theta}_P$, the criterion functional $\phi^D(\xi)$ does depend on $\boldsymbol{\theta}_P$. We have given a condition where $\phi^L(\xi)$ itself is independent of $\boldsymbol{\theta}_P$. Here is another example where we can apply the corollary. Assume that we are interested in $g(\boldsymbol{\theta}) = \theta_p/\theta_{p+q}$, the ratio of the coefficients of the highest orders in the two polynomials P and Q. Consider $g(\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$. When $\boldsymbol{\theta}$ is given, the asymptotic variance of $g(\hat{\boldsymbol{\theta}})$ is $\nabla g^t M^{-1}(\boldsymbol{\theta}, \xi) \nabla g$, where

$$(25) \quad \nabla g = \theta_{p+q}^{-2} (\underbrace{0, \ldots, 0}_{p}, \theta_{p+q}, \underbrace{0, \ldots, 0}_{q-1}, -\theta_p)^t.$$

Thus, for the corresponding Bayesian L-optimal design, $L(\boldsymbol{\theta}) = \nabla g \nabla g^t$. It turns out that

$$(26) \quad \nabla g = \mathbf{B} \frac{1}{\theta_{p+q}^2} (0, \ldots, 0, 1)^t.$$

Clearly, $\mathbf{B}^{-1} L(\boldsymbol{\theta})(\mathbf{B}^{-1})^t$ does not depend on $\boldsymbol{\theta}_P$. In fact, the Bayesian design criterion is

$$\int_\Theta \frac{1}{\theta_{p+q}^4} (0, \ldots, 0, 1) M_0^{-1}(\boldsymbol{\theta}_Q; \xi)(0, \ldots, 0, 1)^t \pi(d\boldsymbol{\theta}),$$

which is indeed independent of $\boldsymbol{\theta}_P$. It is also interesting to note that this L-optimal criterion is a limiting case of the c_e-optimal criterion. Details are given at the end of Section 5.

4. Locally *D*-optimal designs.

4.1. *Existence.* It is known that the locally *D*-optimal design for a simple polynomial regression of degree p has exactly $p+1$ supporting points. The following theorem tells us the number of supporting points for a general rational model (6).

THEOREM 5. *Consider the general rational model* (6). *Assume that* $|\mathbf{B}| \neq 0$. *Then the maximum of* $\log|M(\boldsymbol{\theta}, \xi)|$ *with respect to the set of all probability measures is attained by a measure* ξ^* *which is concentrated on exactly* $p+q+1$ *points with equal weights* $1/(p+q+1)$, *if any one of the following conditions holds*:

   (i) $Q(x)$ *does not vanish for all* $x \in [a, b]$, *and* $\partial^{2(p+q)+1} Q^4(x)/\partial x^{2(p+q)+1}$, *as a function of* x, *has no* 0 *on the open interval* (a, b) (*which implies that* $p < q$);
   (ii) $Q(x)$ *does not vanish for all* $x \in [a, b]$ *and* $p \geq q$;

   (iii) $$\left\{ 1, \frac{1}{Q^4}, \frac{x}{Q^4}, \frac{x^2}{Q^4}, \ldots, \frac{x^{2(p+q)}}{Q^4} \right\}$$

is a T-system on $[a, b]$.

PROOF. By Theorem 2, we have that

$$(27) \quad \log|M(\boldsymbol{\theta}, \xi)| = \log|\mathbf{B}|^2 + \log \det\left[\int_a^b \frac{1}{Q^4} \mathbf{h}_{p+q}(x)\mathbf{h}_{p+q}(x)^t \, d\xi(x) \right].$$

Note that the first term on the right-hand side of the above equation is independent of ξ, and hence can be ignored. Also note that $\log(\cdot)$ is a monotonically increasing function. Therefore, the problem of maximizing $\log|M(\xi)|$ among $\xi \in \Xi$ is equivalent to that of

$$(28) \quad \max_{\xi} \det\left[\int_a^b \frac{1}{Q^4(x)} \mathbf{h}_{p+q}(x)\mathbf{h}_{p+q}(x)^t \, d\xi(x) \right].$$

If we let $w(x) = 1/Q^4(x)$, we conclude that the *D*-optimal design ξ^* is concentrated on exactly $p+q+1$ points by using a theorem in Karlin and Studden (1966), page 33. Furthermore, from Theorem 2, ξ should be equally weighted on its supporting points. $\square$

EXAMPLE 1 (Continued). Again consider the model (19). Assume that $|\mathbf{B}| = \theta_0\theta_2 - \theta_1 \neq 0$. By part (ii) of Theorem 5, the locally *D*-optimal design is equally weighted on three points. If we let $c = d = \theta_2$ in Proposition 1, we find the locally *D*-optimal design ξ^* to be the three-point design with equal weights $1/3$ at $\{a, x^*, b\}$, where

$$(29) \quad x^* = \frac{a + b + 2ab\theta_2}{2 + (a + b)\theta_2}.$$

Note that x^* given by (29) is monotonically decreasing in $\theta_2 \in [0, \infty]$, ranging from the arithmetic mean to the harmonic mean of a and b.

4.2. *Relationship with polynomial regression.* It is known that the D-optimal design for a polynomial regression model is invariant under a change of interval. Suppose we have two linear models:

$$\text{Model 1:} \quad Y = \theta_0 + \theta_1 x + \cdots + \theta_p x^p + \varepsilon, \qquad a \le x \le b,$$

$$\text{Model 2:} \quad Y = \theta_0^* + \theta_1^* z + \cdots + \theta_p^* z^p + \varepsilon, \qquad 0 \le z \le 1,$$

Let $\{x_0, x_1, \ldots, x_p\}$ and $\{z_0, z_1, \ldots, z_p\}$ be the supporting points of the D-optimal designs for model 1 and model 2, respectively. Then we have

$$(30) \qquad\qquad x_j = a + (b - a)z_j, \qquad j = 0, \ldots, p.$$

This result is actually a special case of Lemma 3. For a rational model, the locally D-optimal design depends on the parameter in $Q(x)$. We have the following results.

LEMMA 3. *Suppose that the following two models are used to describe the same experiment using the transformation $x = a + (b - a)z$:*

$$\text{Model 1:} \quad Y = \frac{\theta_0 + \theta_1 x + \cdots + \theta_p x^p}{1 + \theta_{p+1} x + \cdots + \theta_{p+q} x^q} + \varepsilon, \qquad a \le x \le b,$$

$$\text{Model 2:} \quad Y = \frac{\theta_0^* + \theta_1^* z + \cdots + \theta_p^* z^p}{1 + \theta_{p+1}^* z + \cdots + \theta_{p+q}^* z^q} + \varepsilon, \qquad 0 \le z \le 1.$$

Then we have the following one-to-one relationship:

$$(31) \qquad\qquad \theta_j^* = \frac{(b-a)^j}{Q(a)} \sum_{i=j}^{q} \theta_i \binom{i}{j} a^{i-j}, \qquad j = 0, \ldots, p,$$

$$(32) \qquad\qquad \theta_{p+j}^* = \frac{(b-a)^j}{Q(a)} \sum_{i=j}^{q} \theta_{p+i} \binom{i}{j} a^{i-j}, \qquad j = 1, \ldots, q,$$

where $Q(a) = 1 + \sum_{i=1}^{q} \theta_{p+i} a^i$.

PROOF. The result follows from some simple algebra. $\square$

THEOREM 6. *Consider the two models in Lemma 3 for the same experiment. Assume that the locally D-optimal design using model 2 has its support on the k points $z_1, \ldots, z_k$. Then the locally D-optimal design using model 1 with parameters given by (32) has its support on the k points, $x_i = a + (b-a)z_i$, $i = 1, \ldots, k$.*

PROOF. This result follows from simple invariance considerations in linear models. For completeness, we spell out a direct proof. Let $\mathbf{B}$ and $\mathbf{B}^*$ be the

matrices defined by (11) for model 1 and model 2, respectively. For any design ξ under model 1, it follows from (17) that

$$\phi(\xi) = \log(|\mathbf{B}|^2) + \log\left(\det\left\{ \int_a^b \frac{1}{Q^4(x)} \mathbf{h}_{p+q}(x)\mathbf{h}_{p+q}^t(x)\,\xi(dx) \right\} \right)$$

$$= \log(|\mathbf{B}|^2)$$
$$+ \log\left(\det\left\{ \int_0^1 \frac{\mathbf{h}_{p+q}(a+(b-a)z)\mathbf{h}_{p+q}^t(a+(b-a)z)}{Q^4(a+(b-a)z)}\,d\xi(a+(b-a)z) \right\} \right),$$

where $\mathbf{h}_{p+q}(x)$ is defined by (9). It is clear that $Q(a+(b-a)z) = Q(a)Q^*(z)$, where $Q^*(z) = 1 + \theta_{p+1}^* z + \cdots + \theta_{p+1}^* z^{p+q}$. Since $\mathbf{h}_{p+q}(a+(b-a)z)$ is equal to

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ a & b-a & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a^{p+q} & \binom{p+q}{1}(b-a)a^{p+q-1} & \binom{p+q}{2}(b-a)^2 a^{p+q-2} & \cdots & (b-a)^{p+q} \end{pmatrix} \mathbf{h}_{p+q}(z),$$

we then have

$$\phi(\xi) = \log(|\mathbf{B}|^2) - \log(Q^4(a)) + (p+q)(p+q+1)\log(b-a)$$
$$+ \log\left(\det\left\{ \int_0^1 \frac{\mathbf{h}_{p+q}(z)\mathbf{h}_{p+q}^t(z)}{[Q^*(z)]^4}\,d\xi(a+(b-a)z) \right\} \right).$$

Noting that the analog of the design criterion under model 2 is

$$\phi^*(\xi^*) = \log(|\mathbf{B}^*|^2) + \log\left(\det\left\{ \int_0^1 \frac{1}{[Q^*(z)]^4} \mathbf{h}_{p+q}(z)\mathbf{h}_{p+q}^t(z)\,d\xi^*(z) \right\} \right),$$

we obtain the result immediately. $\square$

4.3. *Weak convergence.* In model (6), if $\boldsymbol{\theta}_Q \to \mathbf{0}_q$ the mean response goes to a pth-degree ordinary polynomial regression model, where $\mathbf{0}_q$ is the zero vector in $\mathbb{R}^q$. One might think that the locally D-optimal design for model (6) goes to the locally D-optimal design for a pth-degree ordinary polynomial regression model. This is not quite correct. The answer is a special case of the following general theorem about weak continuity of the locally D-optimal design.

THEOREM 7. *Consider the rational model* (6). *Assume that* $Q(x, \boldsymbol{\theta})$ *does not vanish on* $[a, b]$ *for any* $\boldsymbol{\theta}$ *under consideration. In this case the locally D-optimal design exists. Let* $\xi(\boldsymbol{\theta}_Q)$ *be the locally D-optimal design (which depends only on* $\boldsymbol{\theta}_Q$). *Then* $\xi(\boldsymbol{\theta}_Q)$ *is weakly continuous in* $\boldsymbol{\theta}_Q$ *in the following sense. For any sequence* $\{\boldsymbol{\theta}_Q^m\}_{m\geq 1}$, *assume that* $\boldsymbol{\theta}_Q^m \to \boldsymbol{\theta}_Q^0$ *and the locally D-optimal design* $\xi(\boldsymbol{\theta}_Q^0)$ *is unique. Then* $\xi(\boldsymbol{\theta}_Q^m) \to_w \xi(\boldsymbol{\theta}_Q^0)$.

PROOF.   For the first result, it is enough to prove that every weakly convergent subsequence of $\{\xi(\boldsymbol{\theta}_Q^m)\}_{m \geq 1}$ converges to the same measure $\xi(\boldsymbol{\theta}_Q^0)$. In fact, let $\{\xi(\boldsymbol{\theta}_Q^{m_j})\}_{j \geq 1}$ be any subsequence of $\{\xi(\boldsymbol{\theta}_Q^m)\}_{m \geq 1}$, so that $\{\xi(\boldsymbol{\theta}_Q^{m_j})\} \to_w \xi^*$ for some measure ξ^*. It is easy to show that

$$
\begin{aligned}
\text{tr}&\left[\left(\int_a^b \frac{\mathbf{h}_{p+q}(s)\mathbf{h}_{p+q}^t(s)}{Q^4(s;\boldsymbol{\theta}_Q^{m_j})}\xi(\boldsymbol{\theta}_Q^{m_j}(s;\boldsymbol{\theta}_Q^{m_j}))(ds)\right)^{-1}\left(\frac{\mathbf{h}_{p+q}(x)\mathbf{h}_{p+q}^t(x)}{Q^4(x;\boldsymbol{\theta}_Q^{m_j})}\right)\right] \\
&- (p+q+1) \\
&\to \text{tr}\left[\left(\int_a^b \frac{\mathbf{h}_{p+q}(s)\mathbf{h}_{p+q}^t(s)}{Q^4(s;\boldsymbol{\theta}_Q^0)}\xi^*(ds)\right)^{-1}\left(\frac{\mathbf{h}_{p+q}(x)\mathbf{h}_{p+q}^t(x)}{Q^4(x;\boldsymbol{\theta}_Q^0)}\right)\right] \\
&- (p+q+1)
\end{aligned}
$$

(33)

as $j \to \infty$ for each $x \in [a, b]$. Since each term on the left-hand side is negative for each $x \in [a, b]$, so is the right-hand side. Note that the right-hand side is the directional derivative of ξ^* in the direction ξ_x, so we know that ξ^* is the locally D-optimal design for $\boldsymbol{\theta}_Q^0$, which is unique by assumption. Therefore, $\xi^* = \xi(\boldsymbol{\theta}_Q^0)$. The desired results then follow.   □

COROLLARY 4.   *Under the assumptions of Theorem 7, if $\boldsymbol{\theta}_Q^m \to \mathbf{0}_q$, then $\xi(\boldsymbol{\theta}_Q^m) \to_w \xi_{p+q}$, where $\mathbf{0}_q$ is the zero vector in $\mathbb{R}^q$ and ξ_{p+q} is the D-optimal design for a $(p+q)^{th}$ ordinary polynomial regression model.*

PROOF.   If $\boldsymbol{\theta}_Q^0 = \mathbf{0}_q$, then $\xi(\boldsymbol{\theta}_Q^0)$ is the D-optimal design for the $(p+q)$th-degree polynomial regression model and is indeed unique from Guest (1958).   □

EXAMPLE 1 (Continued).   As $\theta_2 \to 0$, (29) goes to the arithmetic mean of a and b, so that the locally D-optimal design ξ^* tends to the design equally weighted at $\{a, (a+b)/2, b\}$, which is the D-optimal design for the polynomial model $Y = \theta_0 + \theta_1 x + \theta_2 x^2 + \varepsilon$, $x \in \mathscr{X} = [a, b]$. This agrees with the conclusion in Corollary 4.

5. Locally c_e-optimal designs.   The problem of extrapolation with a rational response model follows the result for the ordinary polynomial to some extent. The polynomial case was first derived by Hoel and Levine (1964), who showed that the support of the optimal extrapolation design does not depend on the point of extrapolation x_e. Again consider model (6). From (23), if we define $\mathbf{c} = \mathbf{h}_{p+q}(x_e)/Q^2(x_e)$, then the optimal design ξ minimizing $\phi_e(\xi; x_e)$ is the optimal design minimizing $\mathbf{c}^t M_0^{-1}(\boldsymbol{\theta}_Q, \xi)\mathbf{c}$.

The solution depends on the existence of a certain oscillating polynomial described in the following lemma. Denote a class of functions on $[a, b]$ by

$\mathcal{W} = \{W(x) = \sum_{i=0}^{p+q} a_i x^i / Q^2(x) | a_i \in \mathbb{R}, \ x \in [a, b]\}$ and define the determinant by

$$(34) \quad H\begin{pmatrix} 0, & 1, & \ldots, & p+q \\ x_0, & x_1, & \ldots, & x_{p+q} \end{pmatrix} \equiv \begin{vmatrix} \dfrac{1}{Q^2(x_0)} & \dfrac{1}{Q^2(x_1)} & \cdots & \dfrac{1}{Q^2(x_{p+q})} \\[2mm] \dfrac{x_0}{Q^2(x_0)} & \dfrac{x_1}{Q^2(x_1)} & \cdots & \dfrac{x_{p+q}}{Q^2(x_{p+q})} \\[2mm] \vdots & \vdots & & \vdots \\[2mm] \dfrac{x_0^{p+q}}{Q^2(x_0)} & \dfrac{x_1^{p+q}}{Q^2(x_1)} & \cdots & \dfrac{x_{p+q}^{p+q}}{Q^2(x_{p+q})} \end{vmatrix}.$$

LEMMA 4. *Assume that $Q(x) = Q(x, \theta_Q) \neq 0$ for any $x \in [a, b]$. Then there exists a unique $W^*(x) = \sum_{i=0}^{p+q} a_i^* x^i / Q^2(x) \in \mathcal{W}$ satisfying the following properties*:

(i) $|W^*(x)| \leq 1$;

(ii) *there are $p + q + 1$ points $a \leq s_0 < s_1 < \cdots < s_{p+q} \leq b$, such that $W^*(s_i) = (-1)^{p+q-i}$, $i = 0, 1, \ldots, p+q$. Furthermore, when $U(x) \equiv 1 \in \mathcal{W}$, then $s_0 = a$ and $s_{p+q} = b$, and equality occurs in part (i) only for $x = s_0, s_1, \ldots, s_{p+q}$.*

PROOF. The result follows from Theorem 2.10.1 of Karlin and Studden (1966) and the fact that

$$\frac{1}{Q^2(x)} \mathbf{h}_{p+q}(x) = \frac{1}{Q^2(x)} (1, x, \ldots, x^{p+q})^t$$

forms a T-system on $[a, b]$. □

For $s_0, s_1, \ldots, s_{p+q}$ given by Lemma 4, define the following determinants:

$$(35) \quad D_i = H\begin{pmatrix} 0, & \ldots, & i-1, & i, & \ldots, & p+q-1, & p+q \\ s_0, & \ldots, & s_{i-1}, & s_{i+1}, & \ldots, & s_{p+q}, & x^e \end{pmatrix},$$

where $i = 0, \ldots, p + q$. We know that $D_i > 0$ if $x_e > b$, and the sign of D_i will be $(-1)^{p+q}$ if $x_e < a$. We further let

$$(36) \quad p_i = |D_i| \Big/ \sum_{j=0}^{p+q} |D_j|, \qquad i = 0, \ldots, p+q.$$

Then we have the following theorem for the locally c_e-optimal design.

THEOREM 8. *Assume that $Q(x) \neq 0$ for any $x \in \{x_e\} \cup [a, b]$. Then*:

(a) *For any design ξ,*

$$(37) \quad \phi_e(\xi, x_e) \geq [W^*(x_e)]^2.$$

(b) *Equality occurs in (37) if and only if $\xi = \xi^*$ which puts mass p_i at the point s_i, $i = 0, 1, \ldots, p + q$.*

PROOF. The proof follows the same arguments as in the proof of Theorem 2.2 of Studden (1968) and is therefore omitted. □

REMARK 2. Assume that the conditions of Theorem 8 hold. If $p \geq q$, then:

(a) $Q^2(x)W^*(x)$ is the $(p+q)$th orthogonal polynomial with respect to

$$\frac{1}{\sqrt{(b-x)(x-a)}Q^4(x)};$$

(b) the two boundary supporting points of the c_e-optimal design are $s_0 = a$ and $s_{p+q} = b$;

(c) the middle supporting points $(s_1, \ldots, s_{p+q-1})$ for the c_e-optimal design are the roots of the $(p+q-1)$th orthogonal polynomial $u_{p+q-1}(x)$ with respect to

$$\frac{\sqrt{(b-x)(x-a)}}{Q^4(x)}.$$

PROOF. Part (a) follows from Krein and Nudelman (1977), pages 363 and 364. Part (b) follows from Theorem 2.2 of Studden (1968) and part (c) follows from Lau (1983), page 125. □

We know that the locally c_e-optimal design, say ξ_e, is independent of $\boldsymbol{\theta}_P$, the parameters in $P(x)$. From Lemma 4 and Theorem 8, the $p+q+1$ support points $(s_0, s_1, \ldots, s_{p+q})$ of ξ_e do not depend on x_e (this agrees with the result for c_e-optimal designs under a polynomial regression model). In contrast, the weights at $(s_0, s_1, \ldots, s_{p+q})$ for ξ_e depend on x_e and are not equally weighted. Recall that, in Theorem 5, we showed that the locally D-optimal design, ξ_D say, is equally weighted on $p+q+1$ points. However, we do not know the location of these $p+q+1$ points. In practice, instead of finding ξ_D, one may use the design ξ_e^{ew} which puts equal weight on the $p+q+1$ support points of ξ_e. Such a ξ_e^{ew} is often "close" to ξ_D, as the following example illustrates.

EXAMPLE 1 (Continued). Suppose that we want to extrapolate the experimental dependence to a given point $x_e > b$ for model (19), that is, to estimate, $f(x_e, \boldsymbol{\theta}) = (\theta_0 + \theta_1 x_e)/(1 + \theta_2 x_e)$, the mean response at x_e. Since $p \geq q$, from Remark 2, we know that the locally c_e-optimal design has three supporting points $\{a, s_1, b\}$, where s_1 is the root of the linear function orthogonal to $\sqrt{(b-x)(x-a)}/(1 + \theta_2 x)^4$. Therefore, we have

$$(38) \qquad s_1 = \left\{ \int_a^b \frac{x\sqrt{(b-x)(x-a)}}{(1+\theta_2 x)^4}\,dx \right\} \bigg/ \left\{ \int_a^b \frac{\sqrt{(b-x)(x-a)}}{(1+\theta_2 x)^4}\,dx \right\}.$$

Furthermore, by Theorem 8, the weights for the optimal design at the three points $\{a, s_1, b\}$ are $\{p_0, p_1, p_2\}$, where $p_i = C_i/(C_0 + C_1 + C_2)$, $i = 0, 1, 2$: $C_0 = (1 + \theta_2 a)^2(b - s_1)(x_e - b)(x_e - s_1)$, $C_1 = (1 + \theta_2 s_1)^2(b - a)(x_e - b)(x_e - a)$ and $C_2 = (1 + \theta_2 b)^2(s_1 - a)(x_e - s_1)(x_e - a)$.

TABLE 1
The middle design points x^* and s_1 and the weights (p_1, p_2, p_3)
for the model $Y = (\theta_0 + \theta_1 x)/(1 + \theta_2 x) + \varepsilon,\ x \in [0, 1]$

θ_2	x^*	s_1	x_e	(p_1, p_2, p_3)
0	0.4999999991	0.5000000000	2	$(0.1765, 0.4706, 0.3529)$
			20	$(0.2437, 0.4998, 0.2565)$
			200	$(0.2494, 0.5000, 0.2506)$
1	0.3333333341	0.3333333333	2	$(0.1220, 0.3902, 0.4878)$
			20	$(0.1719, 0.4662, 0.3619)$
			200	$(0.1760, 0.4702, 0.3538)$
10	0.0833333313	0.0833333333	2	$(0.0373, 0.1426, 0.8201)$
			20	$(0.0615, 0.2264, 0.7121)$
			200	$(0.0636, 0.2333, 0.7031)$
100	0.0098039039	0.0098039216	2	$(0.0048, 0.0192, 0.9759)$
			20	$(0.0090, 0.0356, 0.9554)$
			200	$(0.0094, 0.0372, 0.9534)$

Let us compare the numerical values of the two middle design points, x^* [defined by (29) for the locally D-optimal design] and s_1 [defined by (38) for the locally c_e-optimal design]. We chose $a = 0$ and $b = 1$. Then the numerical values of x^* and s_1 for various values of θ_2 are given in Table 1. The weights p_i for the locally c_e-optimal design are also given. It is seen that x^* and s_1 are identical up to 10^{-8}.

Before concluding this section, let us state a relationship between the locally c_e-optimal design and the locally L-optimal design when $L(\boldsymbol{\theta}) = \nabla g^t \nabla g$ and ∇g is given by (25). This arises from estimating θ_p/θ_{p+q}. It is of interest that the supporting points for these two locally optimal designs are identical.

THEOREM 9. *Assume that $Q(x) \neq 0$ for any $x \in [a, \infty)$ and $\theta_{p+q} \neq 0$. Then the supporting points for the locally c_e-optimal design are the same as those of the locally L-optimal design when $L(\boldsymbol{\theta}) = \nabla g^t \nabla g$ and ∇g is given by (25).*

PROOF. From Theorem 8, the c_e-optimal design for estimating $P(x_e)/Q(x_e)$ is independent of x_e. The optimal design for estimating $P(x_e)/Q(x_e)$ is invariant under multiplication by a constant. Let $x_e \to \infty$; then $x_e^{q-p} P(x_e)/Q(x_e) \to \theta_p/\theta_{p+q}$. The result follows immediately. $\square$

6. The Michaelis–Menten model.

6.1. *The model and the D-optimal design.* For many physical and biological phenomena, saturation functions are often described by a rectangular hyperbola. The Michaelis–Menten model

$$(39) \qquad Y = \frac{\theta_1 x}{\theta_2 + x} + \varepsilon, \qquad 0 \le x \le b \le \infty,$$

is frequently used to model the speed of enzyme kinetic reactions, where Y is the reaction velocity, x the concentration of substrate, $\theta_1 > 0$ the maximum velocity of the reaction and $\theta_2 \geq 0$ the half-saturation constant, since the mean response is half-maximal at θ_2. Note that this model is not a special case of model (6).

Currie (1982) concluded that the locally D-optimal design is obtained by taking half the observations at the point θ_2 and the other half as high as possible, namely at b. We will see that this is only approximately true when b is very large. Define

$$(40) \qquad \tilde{\mathbf{B}} = \begin{pmatrix} \theta_2 & 1 \\ -\theta_1 & 0 \end{pmatrix}, \qquad \tilde{\mathbf{h}} \equiv \tilde{\mathbf{h}}(x) = \frac{1}{(\theta_2 + x)^2} \begin{pmatrix} x \\ x^2 \end{pmatrix}.$$

Then $\nabla f(x, \boldsymbol{\theta}) = \tilde{\mathbf{B}}\tilde{\mathbf{h}}(x)$ and $I(\boldsymbol{\theta}, x) = \nabla f \nabla f^t = \tilde{\mathbf{B}}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^t \tilde{\mathbf{B}}^t$.

THEOREM 10.   *Consider model (39).*

(a) *Assume that the marginal prior distribution of θ_2 has one-half mass at each of the points c and d. Among all two-point designs, the Bayesian optimal design ξ puts equal weight on two points, x^* and b, where*

$$(41) \qquad x^* = \frac{\sqrt{cd(b+c)(b+d)} - c}{c+d+b}.$$

(b) *The locally D-optimal design ξ^* is the equally weighted two-point design at $\{x^*, b\}$, where*

$$(42) \qquad x^* = \theta_2 b / (2\theta_2 + b).$$

PROOF.   Part (a) can be proved similarly to Proposition 1. For part (b), by an argument similar to Theorem 5, we can show that the locally D-optimal design is equally supported on two points. These two points can be found from part (a) by letting $c = d = \theta_2$.   □

6.2. *The locally c_e-optimal design.*   Suppose that we want to extrapolate the model to $x_e > b$, that is, to estimate $f(x_e, \boldsymbol{\theta}) = \theta_1 x_e / (\theta_2 + x_e)$. The design criterion is

$$\phi_e(\xi; x_e) = \nabla f(x_e, \boldsymbol{\theta})^t M^{-1}(\boldsymbol{\theta}, \xi) \nabla f(x_e, \boldsymbol{\theta}),$$

$$M(\boldsymbol{\theta}, \xi) = \int_{\mathscr{X}} I(\boldsymbol{\theta}, x)\xi(dx).$$

It is easy to see that

$$\phi_e(\xi; x_e) = \tilde{\mathbf{h}}^t(x_e) M_0^{-1}(\theta_2; \xi)\tilde{\mathbf{h}}(x_e),$$

$$M_0^{-1}(\theta_2; \xi) = \int_{\mathscr{X}} \tilde{\mathbf{h}}(x)\tilde{\mathbf{h}}^t(x)\xi(dx).$$

THEOREM 11.   *The c_e-optimal design at x_e is the two-point design at z^* and b with weights p^* and $1 - p^*$, respectively, where*

$$(43) \qquad z^* = \frac{(\sqrt{2} - 1)\theta_2^2 + \sqrt{2}\theta_2 b}{\theta_2^2 + 4\theta_2 b + 2b^2} b,$$

$$(44) \qquad p^* = \frac{b(x_e - b)(\theta_2 + z^*)^2}{b(x_e - b)(\theta_2 + z^*)^2 + z^*(x_e - z^*)(\theta_2 + b)^2}.$$

PROOF.   This is similar to Theorem 8.   □

The following results give some comparisons for the two locally optimal designs.

PROPOSITION 2.   *For any given θ_2, let x^*, z^* and p^* be given by* (42), (43) *and* (44), *respectively. Then*:

(a) $$\lim_{b \to \infty} x^* = \theta_2;$$

(b) $$\lim_{b \to \infty} z^* = \frac{\sqrt{2}}{2}\theta_2;$$

(c) *the lower locally D-optimal design point is always larger than the optimal design point for extrapolation, that is, $x^* > z^*$, for any $b > 0$;*
(d) *for fixed $b > 0$,*

$$\lim_{x_e \downarrow b} p^* = 0$$

and

$$\lim_{x_e \uparrow \infty} p^* = b\left(\frac{\sqrt{2}}{2} + 1\right)^2 \theta_2^2 \Big/ \left[b\left(\frac{\sqrt{2}}{2} + 1\right)^2 \theta_2^2 + \frac{\sqrt{2}}{2}\theta_2(\theta_2 + b)^2 \right].$$

PROOF.   Parts (a) and (b) are obvious. For part (c), it is enough to check that

$$[(\sqrt{2} - 1)\theta_2 + \sqrt{2}b](b + 2\theta_2) \le 2b^2 + 4b\theta_2 + \theta_2^2.$$

Part (d) follows from part (a) and some algebra.   □

An example of the graph of the supporting points x^* and z^* is shown in Figure 1. For the graph, we chose $\theta_2 = 50$. We see that both x^* and z^* are monotonically increasing in b and satisfy $z^* < x^* < \theta_2$.
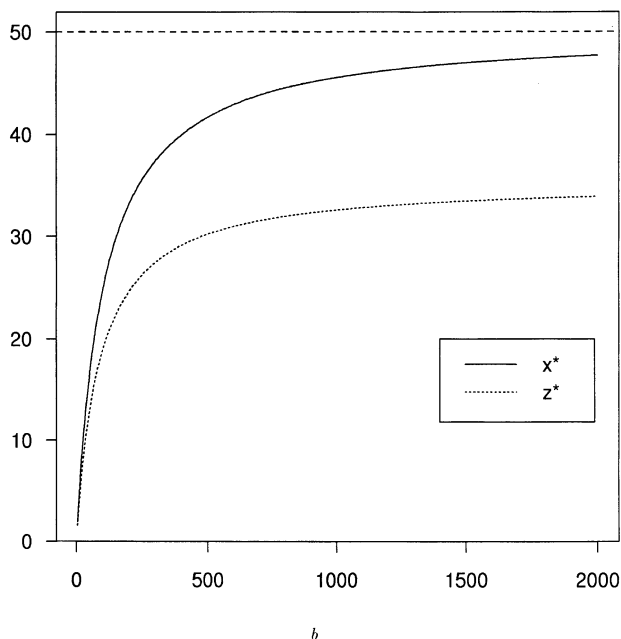
FIG. 1. *Comparison of lower supporting points* $x^* = b\theta_2/(2\theta_2 + b)$ *and* $z^* = b[(\sqrt{2} - 1)\theta_2^2 + \sqrt{2}\theta_2 b]/[\theta_2^2 + 4\theta_2 b + 2b^2]$ *for fixed* $\theta_2 = 50$.

REFERENCES

BOX, G. E. P. and LUCAS, H. L. (1959). Design of experiments in non-linear situations. *Biometrika* **46** 77–90.

CHALONER, K. (1993). A note on optimal Bayesian design for nonlinear problems. *J. Statist. Plann. Inference* **37** 229–235.

CHALONER, K. and LARNTZ, K. (1989). Optimal Bayesian experimental design applied to logistic regression experiments. *J. Statist. Plann. Inference* **21** 191–208.

CHERNOFF, H. (1953). Local optimal designs for estimating parameters. *Ann. Math. Statist.* **24** 586–602.

CURRIE, D. J. (1982). Estimating Michaelis–Menten parameters: bias, variance and experimental design. *Biometrics* **38** 907–919.

DETTE, H. and NEUGEBAUER, H.-M. (1996). Bayesian optimal one point designs for one parameter nonlinear models. *J. Statist. Plann. Inference* **52** 17–31.

FEDOROV, V. V. (1972). *Theory of Optimal Experiment*. Academic Press, New York.

FISHER, R. A. (1922). On the mathematical foundation of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **22** 309–368.

FORD, I., TITTERINGTON, D. M. and KITSOS, C. P. (1989). Recent advances in nonlinear experimental design. *Technometrics* **31** 49–60.

FORD, I., TORSNEY, B. and 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. Ser. B* **54** 569–583.

GUEST, P. G. (1958). The spacing of observations in polynomial regression. *Ann. Math. Statist.* **29** 294–299.

HE, Z. (1993). Optimal designs for rational models. Ph.D. dissertation, Dept. Statistics, Purdue Univ.

HILL, P. D. H. (1980). *D*-optimal designs for partially nonlinear regression models. *Technometrics* **22** 275–276.

HOEL, P. G. and LEVINE, A. (1964). Optimal spacing and weighting in polynomial prediction. *Ann. Math. Statist.* **35** 1553–1560.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Application in Analysis and Statistics.* Interscience, New York.

KATZ, D., AZEN, S. P. and SCHUMITZKY, A. (1981). Bayesian approach to the analysis of nonlinear models: implementation and evaluation. *Biometrics* **37** 137–142.

KREIN, M. G. and NUDELMAN, A. A. (1977). *The Markov Moment Problem And Extremal Problems*, (translated by D. Louvish). Amer. Math. Soc., Providence, RI.

LAU, T. (1983). Theory of canonical moments and its applications in polynomial regression, parts I and II. Technical Reports #83–23 and #83–24, Purdue Univ.

LÄUTER, E. (1974). Experimental design in a class of models. *Math. Operationsforsch. Statist. Ser. Statist.* **5** 379–398.

LINDLEY, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Statist.* **27** 986–1005.

PILZ, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Wiley, New York.

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

SPARROW, P. E. (1979a). Nitrogen response curves of spring barley. *Journal of Agricultural Science, Cambridge* **92** 307–317.

SPARROW, P. E. (1979b). The comparison of five response curves for representing the relationship between the annual dry-matter yield of grass herbage and fertilizer nitrogen. *Journal of Agricultural Science, Cambridge* **93** 513–520.

STUDDEN, W. J. (1968). Optimal designs on Tchebycheff points. *Ann. Math. Statist.* **39** 1435–1447.

TSUTAKAWA, R. K. (1972). Design of an experiment for bioassay. *J. Amer. Statist. Assoc.* **67** 584–590.

WHITTLE, P. (1973). Some general points in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **35** 123–130.

FISHERY AND WILDLIFE RESEARCH CENTER
MISSOURI DEPARTMENT OF CONSERVATION
1110 SOUTH COLLEGE AVENUE
COLUMBIA, MISSOURI 65201

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
MATHEMATICAL SCIENCES BUILDING
WEST LAFAYETTE, INDIANA 47907–1399
E-MAIL: studden@stat.purdue.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF MISSOURI–COLUMBIA
323 MATHEMATICAL SCIENCES BUILDING
COLUMBIA, MISSOURI 65211
E-MAIL: dsun@statserv1.cs.missouri.edu