Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 436239, 15 pages
doi:10.1155/2012/436239

Research Article

Two-Stage Adaptive Optimal Design with Fixed First-Stage Sample Size

Adam Lane and Nancy Flournoy

Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65203, USA

Correspondence should be addressed to Adam Lane, aclpp9@mail.missouri.edu

Received 1 June 2012; Accepted 5 September 2012

Academic Editor: Zhengjia Chen

In adaptive optimal procedures, the design at each stage is an estimate of the optimal design based on all previous data. Asymptotics for regular models with fixed number of stages are straightforward if one assumes the sample size of each stage goes to infinity with the overall sample size. However, it is not uncommon for a small pilot study of fixed size to be followed by a much larger experiment. We study the large sample behavior of such studies. For simplicity, we assume a nonlinear regression model with normal errors. We show that the distribution of the maximum likelihood estimates converges to a scale mixture family of normal random variables. Then, for a one parameter exponential mean function we derive the asymptotic distribution of the maximum likelihood estimate explicitly and present a simulation to compare the characteristics of this asymptotic distribution with some commonly used alternatives.

1. Introduction

Elfving [1] introduced a geometric approach for determining a c-optimal design for linear regression models. Kiefer and Wolfowitz [2] developed the celebrated equivalence theorem which provides an efficient method for verifying if a design is D-optimal, again for a linear model. These two results were generalized by Chernoff [3] and White [4] to include nonlinear models, respectively. See Bartroff [5], O'Brien and Funk [6], and references therein for extensions to the geometric and equivalence approaches. Researchers in optimal design have built an impressive body of theoretical and practical tools for linear models based on these early results. However, advances for nonlinear models have not kept pace.

One reason for the prevalence of the linear assumption in optimal design is that the problem can be explicitly described. The goal of optimal design is to determine precise experiments. Define an approximate design, proposed by Kiefer and Wolfowitz [7], as $\xi = \{x_i, w_i\}_1^K$, where ξ is a probability measure on $\mathcal{X}$ consisting of support points $x_i \in \mathcal{X}$ and corresponding design weights w_i; w_i are rational and defined on the interval $[0,1]$ and $\sum w_i = 1$. Then the optimal design problem is to find the design that maximizes precision for a given experimental interest. Typically, this precision is achieved by maximizing some concave function, ϕ, of Fisher's information matrix. For example, when the estimation of all the parameters is the primary interest then the D-optimality criteria, where ϕ is equal to the determinant of the inverse of Fisher's information, are the most popular method. See Pukelsheim [8] for a detailed discussion of common optimality criteria.

The basic principles for nonlinear models are the same as for linear models except Fisher's information will be a function of the model parameters. As a result, optimal designs depend on the parameters and thus are only optimal in the neighborhood of the true parameters. The term locally optimal design is commonly used for nonlinear optimal designs to reflect this dependence on the parameters of interest.

To overcome this dependence Fisher [9] and Chernoff [3] suggest using expert knowledge to approximate the locally optimal design. Ford et al. [10] suggest optimal designs in nonlinear problems are to be used to provide a benchmark or to construct sequential or adaptive designs. Atkinson et al. [11] suggest using a polynomial expansion to approximate the nonlinear model with a linear one.

Stein [12] provides the earliest two-stage procedure in which the information from the first stage is used to determine design features for the second stage. In this paper we examine a two-stage adaptive optimal design procedure. An adaptive optimal design uses the data from all previous stages to estimate the locally optimal design of the current stage. Many, including Box and Hunter [13], Fedorov [14], White [15], and Silvey [16], have suggested using such designs. Recently, Lane et al. [17], Dragalin et al. [18], Fedorov et al. [19], Yao and Flournoy [20], and so forth have investigated the properties and performance of these procedures.

Lane et al. [17] show that the optimal stage-one sample size is of the order $\sqrt{n}$, where n is the overall sample size, in a two-stage regression model. Luc Pranzato obtains this relationship for a more general model (personal communication, 2012). However, in certain experiments, for example, early phase clinical trials or bioassay studies, it is common to use designs with very small stage-one sample sizes. Current literature has characterized the adaptive optimal design procedure under the assumption that both stage-one and stage-two sample sizes are large.

In this paper we characterize the asymptotic distribution of the maximum likelihood estimate (MLE) when the stage-one sample size is fixed. The distribution for a nonlinear regression model with normal errors and a one parameter exponential mean function is derived explicitly. Then for a specific numeric example the differences between the finite stage-one sample distribution are compared with other candidate approximate distributions.

2. Adaptive Optimal Procedure for a Two-Stage Nonlinear Regression Model with Normal Errors

2.1. The Model

Let $\{y_{ij}\}_{1,1}^{n_i,2}$ be observations from a two-stage experiment, where n_i is the number of observations and x_i is the single-dose level used for the ith stage, $i = 1, 2$. Assume that

$$y_{ij} = \eta(x_i, \theta) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right), \tag{2.1}$$

where $\eta(x,\theta)$ is some nonlinear mean function. In most practical examples it is necessary to consider a bounded design space, that is, $x_i \in \mathcal{X} = [a,b]$, $-\infty < a < b < \infty$. It is assumed that y_{ij} are independent conditional on treatment x_i, where x_1 is fixed and x_2 is selected adaptively. Denote the adaptive design by $\xi_A = \{x_i, w_i\}_1^2$, where $w_i = n_i/n$.

The likelihood for model (2.1) is

$$\mathcal{L}_n(\theta \mid \overline{y}_1, \overline{y}_2) \propto \exp\left\{-\frac{n_1}{2\sigma^2}(\overline{y}_1 - \eta(x_1,\theta))^2 - \frac{n_2}{2\sigma^2}(\overline{y}_2 - \eta(x_2,\theta))^2\right\}, \tag{2.2}$$

where $\overline{y}_i = n_i^{-1}\sum_1^{n_i} y_{ij}$ are the stage specific sample means, and the total score function is

$$S = \frac{d}{d\theta}\ln \mathcal{L}_n(\theta \mid \overline{y}_1, \overline{y}_2) = \frac{n_1}{2\sigma^2}(\overline{y}_1 - \eta(x_1,\theta))\frac{d\eta(x_1,\theta)}{d\theta} + \frac{n_2}{2\sigma^2}(\overline{y}_2 - \eta(x_2,\theta))\frac{d\eta(x_2,\theta)}{d\theta} = S_1 + S_2, \tag{2.3}$$

where S_i represents the score function for the ith stage.

2.2. The Adaptive Optimal Procedure

Fix the first stage design point x_1 and let $\tilde{\theta}_{n_1}$ represent an estimate based on the first-stage complete sufficient statistic $\overline{y}_1$. The locally optimal design point for the second stage is

$$x^* = \arg\max_{x\in\mathcal{X}} \operatorname*{Var}_{\overline{y}_2|x}(S_2) = \arg\max_{x\in\mathcal{X}}\left(\frac{d\eta(x,\theta)}{d\theta}\right)^2, \tag{2.4}$$

which is commonly estimated by $x^*|_{\theta=\tilde{\theta}_{n_1}}$ for use in stage 2. Because the adaptive optimal design literature assumes n_1 is large, the MLE of the second stage design point, $x^*|_{\theta=\widehat{\theta}_{n_1}}$, where $\widehat{\theta}_{n_1}$ is the MLE of θ based on the first stage data, is traditionally used to estimate x^*.

However, when n_1 is small the bias of the MLE can be considerable. Therefore, for some mean functions η using a different estimate would be beneficial. In general, the adaptively selected stage two treatment is

$$x_2 = \arg\max_{x\in\mathcal{X}}\left(\frac{d\eta(x,\theta)}{d\theta}\right)^2\Bigg|_{\theta=\tilde{\theta}_{n_1}}. \tag{2.5}$$

2.3. Fisher's Information

Since $x_1 \in \mathcal{X} = [a, b]$, a bounded design space, but $y \in \mathbb{R}$, there is a positive probability that x_2 will equal a or b. Denote these probabilities as $\pi_a = P(x_2 = a)$ and $\pi_b = P(x_2 = b)$, respectively. Then the per subject information can be written as

$$M(\xi_A, \theta) = \frac{1}{n}\operatorname{Var}(S) = \frac{1}{\sigma^2}\Bigg[w_1\left(\frac{d\eta(x_1,\theta)}{d\theta}\right)^2 + w_2\Bigg(\pi_a\left(\frac{d\eta(a,\theta)}{d\theta}\right)^2 + \pi_b\left(\frac{d\eta(b,\theta)}{d\theta}\right)^2 + E_{x_2}\left[\left(\frac{d\eta(x_2,\theta)}{d\theta}\right)^2 I(a < x_2 < b)\right]\Bigg)\Bigg], \tag{2.6}$$

where x_2 is the random variable defined by the onto transformation (2.5) of $\overline{y}_1$.

3. Asymptotic Properties

We examine three different ways of deriving an asymptotic distribution of the final MLE which may be used for inference at the end of the study. The first is under the assumption that both n_1 and n_2 are large. The second considers the data from the second stage alone. Finally, assume a fixed first-stage sample size and a large second-stage sample size.

3.1. Large Stage-1 and Stage-2 Sample Sizes

If $d\eta(x_2,\theta)/d\theta$ is bounded and continuous and provided common regularity conditions that hold,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_t\right) \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, M^{-1}(\xi^*, \theta)\right), \tag{3.1}$$

as $n_1 \to \infty$ and $n_2 \to \infty$, where $\xi^* = \{(x_1, n_1), (x^*, n_2)\}$. This result is used to justify the common practice of using $x^*|_{\theta=\hat{\theta}_{n_1}}$ to estimate x^* in order to make inferences about θ. However, if $d\eta(x_2,\theta)/d\theta$ is not bounded and continuous then it is very difficult to obtain the result in (3.1) and for certain mean functions the result will not hold. In such cases the asymptotic variance in (3.1) must be replaced with $\lim_{n_1\to\infty} M^{-1}(\xi_A, \theta)$. Lane et al. [17] examine using the exact Fisher's information for an adaptive design ξ_A, $M(\xi_A, \theta)$, instead of $M(\xi^*, \theta)$ in (3.1) to obtain an alternative approximation of the variance of the MLE $\hat{\theta}_n$.

3.2. Distribution of the MLE If Only Second-Stage Data Are Considered

Often pilot data are discarded after being used to design a second experiment then the derivation of the distribution of the MLE using only the second-stage data takes if x_2 to be fixed:

$$\sqrt{n_2}\left(\hat{\theta}_{n_2} - \theta_t\right) \xrightarrow{\mathfrak{D}} \mathcal{N}\left(0, M_2^{-1}(x_2, \theta)\right), \tag{3.2}$$

as $n_2 \to \infty$, where $M_2(x_2,\theta) = \sigma^{-2}(d\eta(x_2,\theta)/d\theta)^2$. The estimate $\widehat{\theta}_{n_2}$ will likely perform poorly in comparison to $\widehat{\theta}_n$ if n_1 and n_2 are relatively of the same size but conceivably may perform quite well when n_1 is much smaller than n. For this reason it represents an informative benchmark distribution.

3.3. Fixed First-Stage Sample Size; Large Second-Stage Sample Size

When the first-stage sample size is fixed and the second stage is large we have the following result.

Theorem 3.1. *For model* (2.1) *with* x_2 *as defined in* (2.5) *if* $d\eta/d\theta \neq 0$ *for all* $x \in \mathcal{X}$, $\theta \in \Theta$, x_2 *is an onto function of* $\overline{y}_1$, $|d\eta/d\theta| < \infty$ *and provided common regularity conditions,*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_t\right) \xrightarrow{\mathfrak{D}} UQ \tag{3.3}$$

as $n_2 \to \infty$, *where* $Q \sim \mathcal{N}(0,\sigma^2)$ *and* $U = ((d\eta(x_2,\theta))/d\theta)^{-1}$ *is a random function of* $\overline{y}_1$.

Proof. As in classical large sample theory (cf. Ferguson [21] and Lehmann [22]):

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \approx \frac{(1/\sqrt{n})S}{-(1/n)(d/d\theta)S}, \tag{3.4}$$

since $S(\widehat{\theta}_n)$ can be expanded around $S(\theta_t)$ as

$$S\left(\widehat{\theta}_n\right) = S(\theta_t) + \left(\widehat{\theta}_n - \theta_t\right)\frac{d}{d\theta}S(\theta_t) + \frac{1}{2}\left(\widehat{\theta}_n - \theta_t\right)^2\frac{d^2}{d\theta^2}S(\theta^*), \tag{3.5}$$

where θ_t is the true value of the parameter and $\theta^* \in (\theta_t, \widehat{\theta}_n)$. Solving for $\sqrt{n}(\widehat{\theta}_n - \theta_t)$ gives

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_t\right) = \frac{(1/\sqrt{n})S(\theta_t)}{-(1/n)\left((d/d\theta)S(\theta_t) + (1/2)\left(\widehat{\theta}_n - \theta_t\right)(d^2/d\theta^2)S(\theta^*)\right)}. \tag{3.6}$$

It can be shown that $\widehat{\theta}_n$ is consistent for θ_t if $n_2 \to \infty$ and $n_1/n \to 0$ which gives the result in (3.4).

Now, decompose the right hand side of (3.4) as

$$\begin{aligned} \frac{(1/\sqrt{n})S}{-(1/n)(d/d\theta)S} &= \frac{(1/\sqrt{n})(S_1 + S_2)}{-(1/n)((d/d\theta)S_1 + (d/d\theta)S_2)} \\ &= \frac{(1/\sqrt{n})S_1}{-(1/n)((d/d\theta)S_1 + (d/d\theta)S_2)} + \frac{(1/\sqrt{n})S_2}{-(1/n)((d/d\theta)S_1 + (d/d\theta)S_2)}. \end{aligned} \tag{3.7}$$

As $n_2 \to \infty$, $S_1/\sqrt{n} \to 0$, $(n_2/n) \to 1$, and $(1/n)(d/d\theta)S_2 \nrightarrow 0$ as $n \to \infty$. Thus, the first term in (3.7) goes to 0 as $n \to \infty$. Write the second term in (3.7) as

$$\frac{(1/\sqrt{n})S_2}{-(1/n)((d/d\theta)S_1 + (d/d\theta)S_2)} = \left(-\frac{(1/n)(d/d\theta)S_1}{(1/\sqrt{n})S_2} - \frac{(1/n)(d/d\theta)S_2}{(1/\sqrt{n})S_2}\right)^{-1}. \tag{3.8}$$

Further as $n_2 \to \infty$, $(1/n)(d/d\theta)S_1 \to 0$ and $(1/\sqrt{n})S_2 \nrightarrow 0$,

$$\begin{aligned} &\frac{(1/n)(d/d\theta)S_1}{(1/\sqrt{n})S_2} \xrightarrow{p} 0, \\ \frac{(1/n)(d/d\theta)S_2}{(1/\sqrt{n})S_2} &= \frac{(1/n)(\overline{y}_2 - \eta(x_2,\theta))(d^2\eta(x_2,\theta)/d\theta^2) + w_2((d\eta(x_2,\theta))/d\theta)^2}{(1/\sqrt{n})n_2(\overline{y}_2 - \eta(x_2,\theta))((d\eta(x_2,\theta))/d\theta)} \\ &= \left(\frac{1}{\sqrt{n}}\right)\frac{d^2\eta(x_2,\theta)/d\theta^2}{d\eta(x_2,\theta)/d\theta} + \frac{\sqrt{w_2}(d\eta(x_2,\theta))/d\theta}{\sqrt{n_2}(\overline{y}_2 - \eta(x_2,\theta))}. \end{aligned} \tag{3.9}$$

The first term in (3.9) goes to 0. To evaluate the second term, it is important to recognize that $\varepsilon_{i2} = \overline{y}_2 - \eta(x_2,\theta) \sim \mathcal{N}(0,\sigma^2/n_2)$ and $\overline{y}_1 \sim \mathcal{N}(0,\sigma^2/n_1)$ are independent and thus

$$\overline{y}_2 - \eta(x_2,\theta), \qquad \frac{d\eta(x_2,\theta)}{d\theta} \tag{3.10}$$

are independent. Because of this independence,

$$\sqrt{n_2}(\overline{y}_2 - \eta(x_2,\theta))\left(\frac{d\eta(x_2,\theta)}{d\theta}\right)^{-1} \sim UQ, \tag{3.11}$$

where U is a random function of $\overline{y}_1$ and $Q \sim \mathcal{N}(0,\sigma^2)$ as determined by $((d\eta(x_2,\theta))/d\theta)^{-1}$. Now, with $\sqrt{w_2} \to 1$ as $n_2 \to \infty$ the result follows from an application of Slutsky's theorem. □

Remark 3.2. Provided $d\eta(x,\theta)/d\theta$ is bounded and continuous UQ is the asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_t)$ as $n \to \infty$. The important case for this exposition is presented in Theorem 3.1. However, the two other potential cases can be shown easily.

Case 1. $n_1 \to \infty$, $n_2 \to \infty$, and $n \to \infty$. As $n_1 \to \infty$, $x_2 \to x^*$ which implies that $U \to ((d(x^*,\theta))/d\theta)^{-1}$, a constant, and thus UQ converges to asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_t)$ given in (3.1).

Case 2. $n_1 \to \infty, n_2$ fixed, and $n \to \infty$. Just as in Case 1, $U \to M^{-1}(x^*,\theta)$, where $M(x^*,\theta) = ((d\eta(x^*,\theta))/d\theta)^{-2}$. Note that $M(x^*,\theta)$ differs from $M(\xi^*,\theta)$ which depends on x_1 and x^*.

Therefore $UQ \to \mathcal{N}(0, \sigma^2 M^{-1}(x^*, \theta))$. Look back at (3.7) in the proof, but now take n_2 to be fixed; $(1/\sqrt{n})S_2 \to 0$ and $(1/\sqrt{n})(d/d\theta)S_2 \to 0$ and the only term left is

$$\frac{(1/\sqrt{n})S_1}{-(1/n)(d/d\theta)S_1}. \tag{3.12}$$

Consider the following: $(1/\sqrt{n})S_1 \to \mathcal{N}(0, \sigma^2 M^{-1}(x^*, \theta))$ and $(1/\sqrt{n})(d/d\theta)S_1 \to M^{-1}(x^*, \theta)$ as $n \to \infty$. Therefore, $\sqrt{n}(\widehat{\theta}_n - \theta_t) \to \mathcal{N}(0, \sigma^2 M^{-1}(x^*, \theta))$ as $n \to \infty$ which is equivalent to UQ.

4. Example: One Parameter Exponential Mean Function

In model (2.1) let $\eta(x, \theta) = e^{-\theta x}$, where $x \in \mathcal{X} = [a, b]$, $0 < a < b < \infty$ and $\theta \in (0, \infty)$. The simplicity of the exponential mean model facilitates our illustration, but it is also important in its own right. For example, Fisher [9] used a variant of this model to examine the information in serial dilutions. Cochran [23] further elaborated on Fisher's application using the same model.

For this illustration we use the MLE of the first-stage data to estimate the second-stage design point. Here,

$$\widehat{\theta}_{n_1} = \begin{cases} \dfrac{-\ln \overline{y}_1}{x}, & \text{if } \overline{y}_1 \in \left(e^{-\overline{\theta} x}, 1\right), \\ 0, & \text{if } \overline{y}_1 \geq 1, \\ \overline{\theta}, & \text{if } \overline{y}_1 \leq e^{-\overline{\theta} x}. \end{cases} \tag{4.1}$$

The adaptively selected second-stage treatment as given by (2.5) is

$$x_2 = \arg\max_{x \in \mathcal{X}} \left(x^2 e^{-2\theta x}\right) = \begin{cases} \widehat{\theta}_{n_1}^{-1}, & \text{if } \overline{y}_1 \in \left(e^{-a^{-1}x_1}, e^{-b^{-1}x_1}\right), \\ b, & \text{if } \overline{y}_1 \geq e^{-b^{-1}x_1}, \\ a, & \text{if } \overline{y}_1 \leq e^{-a^{-1}x_1}. \end{cases} \tag{4.2}$$

Thus, the exact per subject Fisher information is

$$\begin{aligned} M(\xi_A, \theta) = \frac{1}{n}\operatorname{Var}(S) &= w_1 x_1^2 e^{-2\theta x_1} + w_2 \pi_a a^2 e^{-2\theta a} \\ &\quad + w_2 \pi_b b^2 e^{-2\theta b} + w_2 E_{x_2}\left[x_2^2 e^{-2\theta x_2} \cdot I(a < x_2 < b)\right]. \end{aligned} \tag{4.3}$$

For this example $M(\xi_A, \theta) \to M(\xi^*, \theta)$ as $n_1 \to \infty$. For more detailed information on the derivations of (4.1), (4.2), and (4.3) see Lane et al. [17].

The asymptotic distributions of the MLE in Sections 3.1 and 3.2 can be derived easily. For the asymptotic distribution of the MLE in Section 3.3 consider the following corollary. For details on the functions h, v_1, and v_2 see the proof of the corollary.

Corollary 4.1. *If* $\eta(x,\theta) = e^{-\theta x}$ *in model* (2.1) *then*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathfrak{D}} UQ \tag{4.4}$$

as $n \to \infty$, *where* UQ *is defined by*

$$P(UQ \le t) = \begin{cases} P\left(U \ge \dfrac{t}{q} \mid -\infty < q \le 0\right)\Phi(q), & \text{if } t \in (-\infty, 0), \\ P\left(U \le \dfrac{t}{1} \mid 0 < q \le \infty\right)(1 - \Phi(q)), & \text{if } t \in (0, \infty), \end{cases} \tag{4.5}$$

where $\Phi(\cdot)$ *is the standard normal cumulative distribution function. Let* $\Psi(q) = \Phi(\sqrt{n}(q - \eta(x,\theta))/\sigma)$ *and* $h(s) = s^{-1}e^{\theta s}$. *Then if* $h(a) < h(b)$,

$$\begin{aligned} P\left(U \ge \frac{t}{q} \mid -\infty < q \le 0\right)\Phi(q) = {} & \Phi\left(\frac{t}{\sigma h(1/\theta)}\right) \\ & + [1 - (\Psi(v_2(h(a))) - \Psi(v_1(h(a))))] \\ & \times \left[\Phi\left(\frac{t}{\sigma h(a)}\right) - \Phi\left(\frac{t}{\sigma h(1/\theta)}\right)\right] \\ & + [\Psi(v_2(h(b))) - \Psi(v_2(h(a)))] \\ & \times \left[\Phi\left(\frac{t}{\sigma h(b)}\right) - \Phi\left(\frac{t}{\sigma h(a)}\right)\right], \\ P\left(U \le \frac{t}{q} \mid 0 < q \le \infty\right)(1 - \Phi(q)) = {} & \Phi\left(\frac{t}{\sigma h(b)}\right) \\ & + [\Psi(v_2(h(a))) - \Psi(v_1(h(a)))] \\ & \times \left[\Phi\left(\frac{t}{\sigma h(1/\theta)}\right) - \Phi\left(\frac{t}{\sigma h(a)}\right)\right] \\ & + [1 - (\Psi(v_2(h(b))) - \Psi(v_2(h(a))))] \\ & \times \left[\Phi\left(\frac{t}{\sigma h(a)}\right) - \Phi\left(\frac{t}{\sigma h(b)}\right)\right]. \end{aligned} \tag{4.6}$$

If $h(b) < h(a)$, *then*

$$\begin{aligned} P\left(U \ge \frac{t}{q} \mid -\infty < q \le 0\right)\Phi(q) = {} & \Phi\left(\frac{t}{\sigma h(1/\theta)}\right) \\ & + [1 - (\Psi(v_2(h(b))) - \Psi(v_1(h(b))))] \\ & \times \left[\Phi\left(\frac{t}{\sigma h(b)}\right) - \Phi\left(\frac{t}{\sigma h(1/\theta)}\right)\right] \\ & + [\Psi(v_1(h(b))) - \Psi(v_1(h(a)))] \\ & \times \left[\Phi\left(\frac{t}{\sigma h(b)}\right) - \Phi\left(\frac{t}{\sigma h(a)}\right)\right], \end{aligned}$$

$$P\left(U \le \frac{t}{q} \mid 0 < q \le \infty\right)(1 - \Phi(q)) = \Phi\left(\frac{t}{\sigma h(a)}\right)$$
$$+ [\Psi(v_2(h(b))) - \Psi(v_1(h(b)))]$$
$$\times \left[\Phi\left(\frac{t}{\sigma h(1/\theta)}\right) - \Phi\left(\frac{t}{\sigma h(b)}\right)\right]$$
$$+ [1 - (\Psi(v_1(h(b))) - \Psi(v_1(h(a))))]$$
$$\times \left[\Phi\left(\frac{t}{\sigma h(b)}\right) - \Phi\left(\frac{t}{\sigma h(a)}\right)\right]. \tag{4.7}$$

Proof. First, we find the distribution of U where $U = h(z)$ and the random variable z is defined by

$$z = \begin{cases} -\dfrac{x_1}{\ln \overline{y}_1}, & \text{if } \overline{y}_1 \in \left(e^{-x_1/a}, e^{-x_1/b}\right), \\ -\dfrac{x_1}{\ln a}, & \text{if } \overline{y}_1 \le e^{-x_1/a}, \\ -\dfrac{x_1}{\ln b}, & \text{if } \overline{y}_1 \ge e^{-x_1/b}. \end{cases} \tag{4.8}$$

Figure 1 illustrates the map from U to $z \in [a, b]$ where $\theta = 1$, $\sigma = .5$, $a = .25$, and $b = 4$.

Lambert's product log function (cf. Corless et al. [24]) is defined as the solutions to

$$we^{w} = c \tag{4.9}$$

for some constant c. Denote the solutions to (4.9) by $W(w)$. Let

$$V(c) = \arg_{\overline{y}_1}\left\{\left(-\frac{x_1}{\ln \overline{y}_1}\right)^{-1} \exp\left\{\theta \frac{-x_1}{\ln \overline{y}_1}\right\} = c\right\}. \tag{4.10}$$

Then

$$V(c) = \exp\left\{\frac{\theta x_1}{W(-\theta/c)}\right\}. \tag{4.11}$$

The W function is real valued on $w \ge -1/e$, single valued at $w = -1/e$, and double valued on $w \in (-1/e, 0)$. $U \in \{\theta e, \max\{h(a), h(b)\}\}$, $x_1 \in [a, b]$, $0 < a < b < \infty$. Therefore $V(c)$ is real valued for all $\theta \in (0, \infty)$. For simplicity, define $v_1 = \min V(c)$ and $v_2 = \max V(c)$ for a given c.

We present the proof for the cumulative distribution function (CDF) of U and the CDF of UQ for the case where $x^* \in [a, b]$ and $h(a) < h(b)$. The derivation of the distributions under alternative cases is tedious and does not differ greatly from this case.

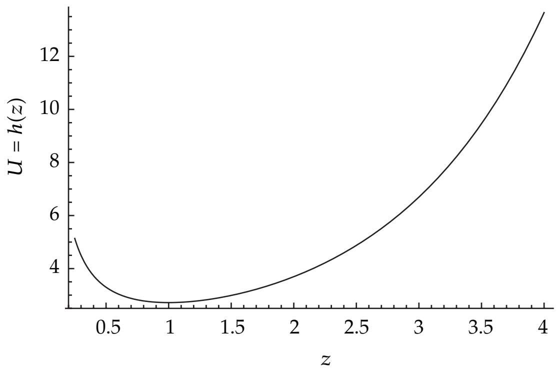

Figure 1: Map of $z = -x_1/\ln \overline{y}_1$ for $\theta = 1$, $a = .25$, and $b = 4$.

Note in this case the domain of U is $[h(1/\theta) = \theta e, h(b)]$. If $h(1/\theta) < U < h(a)$, then

$$P(U \leq t_1) = P\left(h\left(\overline{y}_1\right) < t\right) = P\left(v_1(t_1) < \overline{y}_1 < v_2(t_1)\right) = \Psi(v_2(t_1)) - \Psi(v_1(t_1)). \tag{4.12}$$

If $U = h(a)$, then

$$P(U \leq h(a)) = \Psi(v_2(h(a))). \tag{4.13}$$

If $U \in (h(a), h(b))$, then

$$P(U \leq t_1) = P\left(v_1(t_1) < \overline{y}_1 < v_2(t_1)\right). \tag{4.14}$$

However, since $t_1 < h(a)$ $P(\overline{y}_1 < v_1(t_1)) = 0$,

$$P(U \leq t_1) = \Psi(v_2(t_1)). \tag{4.15}$$

If $U \geq h(b)$, then

$$P(U \leq h(b)) = 1. \tag{4.16}$$

Thus,

$$P(U \leq t_1) = \begin{cases} 0, & \text{if } t_1 \leq h\left(\dfrac{1}{\theta}\right), \\ \Psi(v_2(t_1)) - \Psi(v_2(t_1)), & \text{if } t_1 \in \left(h\left(\dfrac{1}{\theta}\right), h(a)\right), \\ \Psi\left(e^{-x_1/a}\right), & \text{if } t_1 = h(a), \\ \Psi(v_2(t_1)), & \text{if } t_1 \in (h(a), h(b)), \\ 1, & \text{if } h(b) \leq t_1 \leq \infty. \end{cases} \tag{4.17}$$

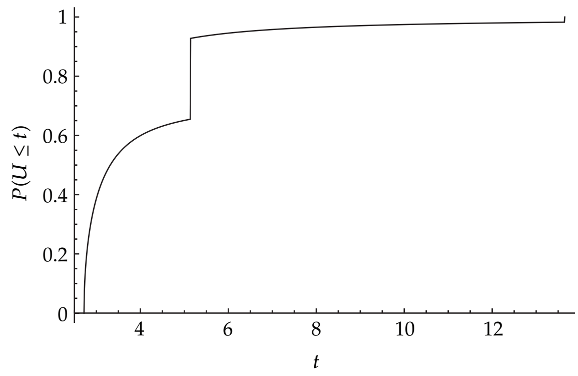

Figure 2: CDF of U for $\theta = 1$, $x_1 = 2$, $n_1 = 5$, $\sigma = .5$, $a = .25$, and $b = 4$.

Figure 2 plots the CDF of U for $\theta = 1$, $x_1 = 2$, $n_1 = 5$, $\sigma = .5$, $a = .25$, and $b = 4$. The distribution is a piecewise function with discontinuities at the boundary points a and b.

Now consider the distribution of UQ. Recall $q \sim \mathcal{N}(0, \sigma^2)$ and U and Q are independent. If $t \in (-\infty, 0)$, then

$$\begin{aligned} P(UQ \le t) = {} & P\left(U \ge \frac{t}{q} \mid 0 < \frac{t}{q} \le h\left(\frac{1}{\theta}\right)\right) P\left(0 < \frac{t}{q} \le h\left(\frac{1}{\theta}\right)\right) \\ & + P\left(U \ge \frac{t}{q} \mid h\left(\frac{1}{\theta}\right) < \frac{t}{q} \le h(a)\right) P\left(h\left(\frac{1}{\theta}\right) < \frac{t}{q} \le h(a)\right) \\ & + P\left(U = h(a) \mid \frac{t}{q} = h(a)\right) P\left(\frac{t}{q} = h(a)\right) \\ & + P\left(U \ge \frac{t}{q} \mid h(a) < \frac{t}{q} \le h(b)\right) P\left(h(a) < \frac{t}{q} \le h(b)\right) \\ & + P\left(U = h(b) \mid \frac{t}{q} = h(b)\right) P\left(\frac{t}{q} = h(b)\right) \\ & + P\left(U \ge \frac{t}{q} \mid h(b) < \frac{t}{q} \le \infty\right) P\left(h(b) < \frac{t}{q} < \infty\right). \end{aligned} \tag{4.18}$$

The distribution is symmetric, thus the derivation of the CDF if $t \in (0, \infty)$ is analogous. □

4.1. Comparisons of Asymptotic Distributions

First, consider the distribution described in (3.1) using $M(\xi_A, \theta)$ in place of $M(\xi^*, \theta)$ and the distribution described in (3.2). When n_1 is significantly smaller than n_2, $M(\xi_A, \theta)$ and $M(x_2, \theta)$ can differ significantly as a function of $\overline{y}_1$. This is primarily because $M(x_2, \theta)$ is a function of x_2, whereas $M(\xi_A, \theta)$ is an average over $\overline{y}_1$. Through simulation it can be seen that a $\mathcal{N}(0, M^{-1}(x_2, \theta))$ is a better approximate distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ than $\mathcal{N}(0, M^{-1}(\xi_A, \theta))$ for only a small interval of x_2, and this interval has a very small probability. For these reasons the distribution of the MLE using only the second stage data as described in Section 3.2 is not considered further.

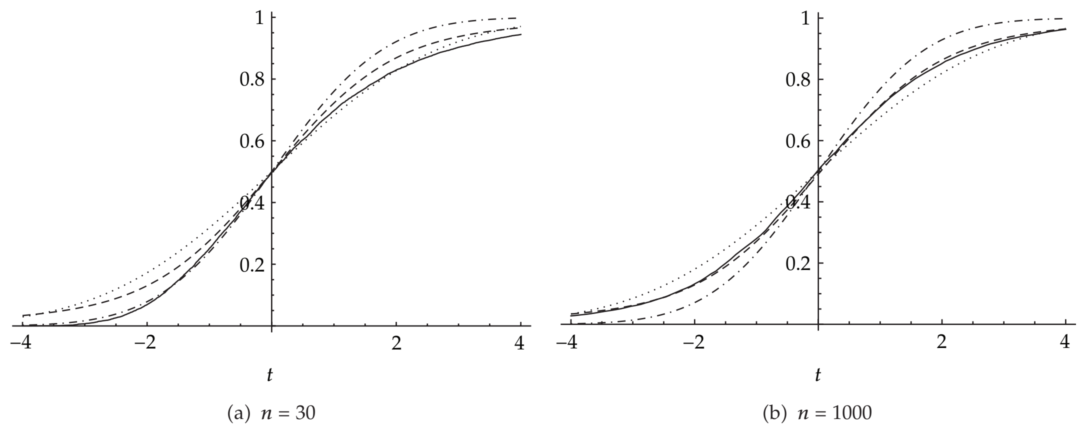

Figure 3: In each plot the solid line represents the CDF of $\sqrt{n}(\widehat{\theta}_n - \theta)$ obtained via Monte Carlo simulations. The dotted-dashed line is the $P(T_1 \leq t)$, where $T_1 \sim \mathcal{N}(0, M^{-1}(\xi^*, \theta))$. The dotted line is the $P(T_2 \leq t)$, where $T_2 \sim \mathcal{N}(0, M^{-1}(\xi_A, \theta))$. The dashed line is the $P(T_3 \leq t)$, where $T_3 \sim Q$. Values $\theta = 1$, $x_1 = 2$, $n_1 = 5$, $\sigma = .5$, $a = .25$, and $b = 4$ were used.

Now for a set of numeric examples consider three distributions: (3.1), (3.1) using $M(\xi_A, \theta)$ in place of $M(\xi^*, \theta)$ and the distribution of UQ defined in (3.3). An asymptotic distribution can be justified in inference if it is approximately equal to the true distribution. In this case the true distribution is that of $\sqrt{n}(\widehat{\theta}_n - \theta)$. However, $\widehat{\theta}_n$ does not have a closed form and thus its distribution cannot be obtained analytically or numerically. To approximate this distribution 10,000 Monte Carlo simulations have been completed for each example to create a benchmark distribution.

Figure 3 plots the three different candidate approximate distributions, found exactly using numerical methods, together with the distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ approximated using Monte Carlo simulations, for $\theta = 1$, x_1, $\sigma = .5$, $a = .25$, $b = 4$, $n_1 = 5$, and $n = \{30, 1000\}$. Note the y-axis represents $P(T_i \leq t)$, $i = 1, 2, 3$, where T_1 is $\mathcal{N}(0, M^{-1}(\xi^*, \theta))$, T_2 is $\mathcal{N}(0, M^{-1}(\xi_A, \theta))$, and T_3 is UQ. When $n = 30$ it is difficult, graphically, to determine if T_2 or T_3 provides a better approximation for $\sqrt{n}(\widehat{\theta}_n - \theta)$. It seems that if $t \in (-4, 0)$ the distribution T_3 is preferable to T_2; however, when $t \in (0, 4)$ the opposite appears to be the case. It is fairly clear that for this example T_1 performs poorly.

When $n = 1000$, it is clear that T_3 is much closer to $\sqrt{n}(\widehat{\theta}_n - \theta)$ than both T_1 and T_2. Further, comparing the two plots one can see how the distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ has nearly converged to UQ but still differs from those T_1 and T_2 significantly, as predicted by Theorem 3.1 and Corollary 4.1.

Using only graphics it is difficult to assess which of T_1, T_2, and T_3 is nearest $\sqrt{n}(\widehat{\theta}_n - \theta)$ for a variety of cases. To get a better understanding, the integrated absolute difference of the CDFs of T_1, T_2, and T_3 versus that of $\sqrt{n}(\widehat{\theta}_n - \theta)$ for $x_1 = 2$, $\sigma = .5$, $a = .25$, $b = 4$, $n = \{5, 10, 15\}$, and $n = \{30, 50, 100, 400\}$ is presented in Table 1. First, consider the table where $\theta = .5$. The locally optimal stage-1 design point is $x_1 = 2$ when $\theta = .5$; as a result this scenario is the most generous to distribution T_1. However, even for this ideal scenario T_3 outperforms T_1 and T_2 for all values of n_1. In many cases the difference between T_3 and T_1 is quite severe. In this scenario T_3 outperforms T_2; however, the differences are not great.

Table 1: Integrated absolute difference of the cumulative distributions (×100) of $T_1 \sim \mathcal{N}(0, M^{-1}(\xi^*, \theta))$, $T_2 \sim \mathcal{N}(0, M^{-1}(\xi_A, \theta))$, and $T_3 \sim UQ$ versus the approximate cumulative distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ obtained via Monte Carlo simulations for various n_1 and various moderate sizes of n. The values $\theta = 1$, $x_1 = 2$, $\sigma = .5$, $a = .25$, and $b = 4$ were used.

(a) ($\theta = .5$)

n_1	$n = 30$			$n = 50$			$n = 100$			$n = 400$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	19	24	11	16	17	8	15	16	7	14	15	5
10	11	13	8	9	12	7	9	11	6	7	12	4
15	9	10	8	8	9	6	6	9	5	4	10	3

(b) ($\theta = 1$)

n_1	$n = 30$			$n = 50$			$n = 100$			$n = 400$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	30	33	25	30	27	19	34	24	12	39	21	6
10	40	40	32	26	27	22	23	28	16	26	20	8
15	34	34	33	27	28	24	21	23	17	18	20	9

(c) ($\theta = 1.5$)

n_1	$n = 30$			$n = 50$			$n = 100$			$n = 400$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
5	32	33	31	39	25	25	42	21	23	42	17	21
10	34	33	22	27	25	16	32	22	10	35	19	12
15	35	35	32	26	26	21	26	22	13	28	21	7

Next, examine the results for $\theta = 1$ and $\theta = 1.5$. Once again T_3 outperforms T_1 and T_2 in all but 2 cases, where in many cases its advantage is quite significant. Also note that T_2 outperforms T_1 about half the time when $\theta = 1$ and the majority of the time when $\theta = 1.5$. This supports our observation that when the distance between x_1 and x^* increases the performance of T_1 compared with T_2 and T_3 worsens which indicates a lack of robustness for the commonly used distribution T_1. This lack of robustness is not evident for T_1 and T_2.

One final comparison is motivated by the fact that if $n_1 \to \infty$, T_1, T_2, and T_3 have the same asymptotic distribution. Although our method is motivated by the scenario where n_1 is a small pilot study, there is no theoretical reason that T_3 will not perform competitively when n_1 is large. Table 2 presents the integrated differences for the distributions T_2 and T_3 from $\sqrt{n}(\widehat{\theta}_n - \theta)$ for $x_1 = 2$, $\theta = 1$, $\sigma = .5$, $a = .25$, $b = 4$, $n_1 = \{50, 100, 200\}$, and $n = \{400, 1000\}$. T_1 is not included in the table due to the lack of robustness; it can perform better or worse than the other two distributions based on the value of θ. Even with larger values of n_1, T_3 performs slightly better when $n_1 = 50$ and 100 and only slightly worse when $n = 200$ indicating that using T_3 is robust for moderately large n_1.

5. Discussion

Assuming a finite first-stage sample size and a large second-stage sample size, we have shown for a general nonlinear one parameter regression model with normal errors that the asymptotic distribution of the MLE is a scale mixture distribution. We considered only one parameter for simplicity and clarity of exposition.

Table 2: Integrated absolute difference of the cumulative distributions (×100) of $T_1 \sim \mathcal{N}(0, M^{-1}(\xi^*, \theta))$, $T_2 \sim \mathcal{N}(0, M^{-1}(\xi_A, \theta))$, and $T_3 \sim UQ$ versus the approximate cumulative distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ obtained via Monte Carlo simulations for various n_1 and various large sizes of n. The values $\theta = 1$, $x_1 = 2$, $\sigma = .5$, $a = .25$, and $b = 4$ were used.

n_1	$n = 400$		$n = 1000$	
	T_2	T_3	T_2	T_3
50	13	9	13	5
100	10	9	8	4
200	11	14	4	7

For the one parameter exponential mean function, the distribution of the adaptively selected second-stage treatment and the asymptotic distribution of the MLE were derived assuming a finite first-stage sample size and a large second-stage sample size. Then the performance of the normalized asymptotic distribution of the MLE, UQ, was analyzed and compared to popular alternatives for a set of simulations.

The distribution of UQ was shown to represent a considerable improvement over the other proposed distributions when n_1 was considerably smaller than n. This was true even when n_1 is moderately large in size.

Since the optimal choice of n_1 was shown to be of the order $\sqrt{n}$ for this model in Lane et al. [17], the usefulness of these findings could have significant implications for many combinations of n_1 and n.

Suppose it is desired that $P(D_1 \leq \sqrt{n}(\widehat{\theta} n - \theta) \leq D_2 = 1 - \alpha)$, where α is the desired confidence level and θ_t is the true parameter. If one was to use the large sample approximate distribution given in (3.1), D_1 and D_2, and therefore n, cannot be determined until after stage 1. However, using (3.1) with $M(\xi_A, \theta)$ in place of $M(\xi^*, \theta)$ or by using UQ on can compute the overall sample size necessary to solve for D_1 and D_2 before stage one is initiated. One could determine n initially using (3.1) with $M(\xi_A, \theta)$ or UQ and then update this calculation after stage-1 data is available. Such same size recalculation requires additional theoretical justification and investigation of their practical usefulness.

We have not, in this paper, addressed the efficiency of the estimate $\widehat{\theta}_n$. One additional way to improve inference would be to find biased adjusted estimates $\widetilde{\theta}_n$ that are superior to $\widehat{\theta}_n$ for finite samples. We have not investigated the impact on inference of estimating the variances in the distributions of UQ, $\mathcal{N}(0, M^{-1}(\xi^*, \theta))$, $\mathcal{N}(0, M^{-1}(\xi_A, \theta)$, and $\mathcal{N}(0, M^{-1}(x_2, \theta))$. Instead, the distributions themselves are compared. For some details on the question of estimation and consistency see Lane et al. [17] and Yao and Flournoy [20].

Acknowledgment

The authors would like to thank the reviewers for their helpful comments and suggestions.

References

[1] G. Elfving, "Optimum allocation in linear regression theory," *The Annals of Mathematical Statistics*, vol. 23, pp. 255–262, 1952.

[2] J. Kiefer and J. Wolfowitz, "The equivalence of two extremum problems," *Canadian Journal of Mathematics*, vol. 12, pp. 363–366, 1960.

[3] H. Chernoff, "Locally optimal designs for estimating parameters," *The Annals of Mathematical Statistics*, vol. 24, no. 4, pp. 586–602, 1953.

[4] L. V. White, "An extension of the general equivalence theorem to nonlinear models," *Biometrika*, vol. 60, pp. 345–348, 1973.
[5] J. Bartroff, "A new characterization of Elfving's method for high dimensional computation," *Journal of Statistical Planning and Inference*, vol. 142, no. 4, pp. 863–871, 2012.
[6] T. E. O'Brien and G. M. Funk, "A gentle introduction to optimal design for regression models," *The American Statistician*, vol. 57, no. 4, pp. 265–267, 2003.
[7] J. Kiefer and J. Wolfowitz, "Optimum designs in regression problems," *The Annals of Mathematical Statistics*, vol. 30, no. 2, pp. 271–294, 1959.
[8] F. Pukelsheim, *Optimal Design of Experiments*, vol. 50 of *Classics in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 2006.
[9] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh, UK, 1947.
[10] I. Ford, B. Torsney, and C. F. J. Wu, "The use of a canonical form in the construction of locally optimal designs for nonlinear problems," *Journal of the Royal Statistical Society B*, vol. 54, no. 2, pp. 569–583, 1992.
[11] A. C. Atkinson, A. N. Donev, and R. D. Tobias, *Optimum Experimental Designs with SAS*, vol. 34 of *Oxford Statistical Science Series*, Oxford University Press, Oxford, UK, 2007.
[12] C. Stein, "A two-sample test for a linear hypothesis whose power is independent of the variance," *The Annals of Mathematical Statistics*, vol. 16, no. 3, pp. 243–258, 1945.
[13] G. E. P. Box and W. G. Hunter, "Sequential design of experiments for nonlinear models," in *Proceedings of the Scientific Computing Symposium: Statistics*, J. J. Korth, Ed., pp. 113–137, IBM, White Plains, NY, USA, 1965.
[14] V. V. Fedorov, *Theory of Optimal Experiments*, Academic Press, New York, NY, USA, 1972.
[15] L. V. White, *The optimal design of experiments for estimation of nonlinear models [Ph.D. thesis]*, University of London, London, UK, 1975.
[16] S. D. Silvey, *Optimal Design: An Introduction to the Theory for Parameter Estimation*, Chapman & Hall, London, UK, 1980.
[17] A. Lane, P. Yao, and N. Flournoy, Information in a Two-Stage Adaptive Optimal Design, Isaac Newton Institute for the Mathematical Sciences Preprint Series, Cambridge, UK, 2012, http://www.newton.ac.uk/preprints/NI12012.pdf.
[18] V. Dragalin, V. Fedorov, and Y. Wu, "Adaptive designs for selecting drug combinations based on efficacy-toxicity response," *Journal of Statistical Planning and Inference*, vol. 138, no. 2, pp. 352–373, 2008.
[19] V. Fedorov, N. Flournoy, Y. Wu, and R. Zhang, *Best Intention Designs in Dose-Finding Studies*, Isaac Newton Institute Preprint Series, Isaac Newton Institute for Mathematical Sciences, 2011, http://www.newton.ac.uk/preprints.html.
[20] P. Yao and N. Flournoy, "Information in a two-stage adaptive optimal design for normal random variables having a one parameter exponential mean function," in *MoDa 9—Advances in Model-Oriented Design and Analysis*, A. Giovagnoli, A. C. Atkinson, B. Torsney, and C. May, Eds., pp. 229–236, Springer, New York, NY, USA, 2010.
[21] T. S. Ferguson, *A Course in Large Sample Theory*, Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, Fla, USA, 1996.
[22] E. L. Lehmann, *Elements of Large-Sample Theory*, Springer Texts in Statistics, Springer, New York, NY, USA, 1999.
[23] W. Cochran, "Experiments for nonlinear functions," *The Journal of the American Statistical Association*, vol. 68, no. 344, pp. 771–781, 1973.
[24] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert *W* function," *Advances in Computational Mathematics*, vol. 5, no. 4, pp. 329–359, 1996.


Advances in Operations Research
Advances in Decision Sciences
Journal of Applied Mathematics
Algebra
Journal of Probability and Statistics
The Scientific World Journal
International Journal of Differential Equations
Hindawi
Submit your manuscripts at
http://www.hindawi.com
International Journal of Combinatorics
Advances in Mathematical Physics
Journal of Complex Analysis
Journal of Mathematics
Mathematical Problems in Engineering
Abstract and Applied Analysis
Discrete Dynamics in Nature and Society
International Journal of Mathematics and Mathematical Sciences
Journal of Discrete Mathematics
Journal of Function Spaces
International Journal of Stochastic Analysis
Journal of Optimization