Nonlinear Experiments: Optimal Design and Inference Based on Likelihood

PROBAL CHAUDHURI and PER A. MYKLAND*

Nonlinear experiments involve response and regressors that are connected through a nonlinear regression-type structure. Examples of nonlinear models include standard nonlinear regression, logistic regression, probit regression. Poisson regression, gamma regression, inverse Gaussian regression, and so on. The Fisher information associated with a nonlinear experiment is typically a complex nonlinear function of the unknown parameter of interest. As a result, we face an awkward situation. Designing an efficient experiment will require knowledge of the parameter, but the purpose of the experiment is to generate data to yield parameter estimates! Our principal objective here is to investigate proper designing of nonlinear experiments that will let us construct efficient estimates of parameters. We focus our attention on a very general nonlinear setup that includes many models commonly encountered in practice. The experiments considered have two fundamental stages: a static design in the initial stage, followed by a fully adaptive sequential stage in which the design points are chosen sequentially, exploiting a *D*-optimality criterion and using parameter estimates based on available data. We explore the behavior of the maximum likelihood estimate when observations are generated from such an experiment. Two major technical hurdles are (1) the dependent nature of the data obtained from an adaptive sequential experiment and (2) the randomness in the total Fisher information associated with the experiment. Our analysis exploits a martingale structure rooted in the likelihood. We derive sufficient conditions that will ensure convergence of the chosen design to a *D*-optimal one as the number of trials grows. Besides ensuring the large sample optimality of the design, the convergence of the average Fisher information provides an ergodicity condition related to the growth of the martingale processes intrinsically associated with the likelihood. This key observation eventually yields the first-order efficiency of the maximum likelihood estimate via martingale central limit theorem and confirms the asymptotic validity of statistical inference based on the likelihood.

KEY WORDS: Adaptive sequential design; *D*-optimality; Fisher information; Generalized linear models; Martingales; Maximum likelihood estimation; Nonlinear regression.

1. NONLINEAR EXPERIMENTS AND OPTIMAL DESIGNS

Consider a controlled experiment with a response variable Y and a deterministic regressor X that has values chosen by the experimenter from an experiment space Ω. In the case of a factorial experiment with each factor having a finite number of levels, Ω is a finite set. On the other hand, for experiments involving continuous regressors, Ω will be an interval or a subset (e.g., a rectangle) of an Euclidean space. A combination of these two cases can be encountered in analysis of covariance-type problems, where some of the regressors are discrete and some are continuous. The outcome Y of the experiment is subject to random variation, and we will denote the response space (i.e., the set of all possible values of Y) by $\mathcal{R}$. $\mathcal{R}$ will be finite (e.g., experiments with binary or polytomous response) or countably infinite (e.g., Poisson experiment) for an experiment that generates discrete response, and it will be an interval on the real line when the response is a continuous random variable. We will assume that the conditional distribution of Y given $X = x$ has a probability mass function (if Y is discrete) or a probability density function (if Y is continuous), which we will denote by $f(y \mid \theta, \phi, x)$. The form of $f(y \mid \theta, \phi, x)$ will be completely known except for the values of θ and ϕ. Here $\theta \in \Theta$ is an Euclidean parameter, which will be our parameter of principal interest, and ϕ, another Euclidean parameter, will be considered a nuisance parameter. Some examples of commonly used nonlinear models follow.

Example 1.1. The well-known Michaelis–Menten model used in enzyme kinetics connects the observed velocity Y of a reaction with the substrate concentration X through the equation $Y = \alpha X(\beta + X)^{-1} + e$, where e is a random variable with zero mean and $\theta = (\alpha, \beta)$ is the unknown parameter of interest. This model and its various extensions have been widely used in studying the dynamics of various chemical and biochemical reactions (see Bates and Watts 1988; Seber and Wild 1989).

Example 1.2. In irreversible chemical reactions, the concentration Y of an intermediate substance at time X is sometimes assumed to satisfy the nonlinear regression structure $Y = \alpha(\alpha - \beta)^{-1}\{\exp(-\beta X) - \exp(-\alpha X)\} + e$. As in the preceding example, here also e is assumed to be a random variable with zero mean, and $\theta = (\alpha, \beta)$ is our parameter of interest. This model was considered by Box and Lucas (1959), Hill and Hunter (1974), and others (see Seber and Wild 1989).

Both of the preceding examples and many other nonlinear models arising in various scientific disciplines will fit into a general nonlinear regression setup in which the influence of X on Y is modeled as $Y = g(\theta, X) + e$. Here g is a real-valued nonlinear function with a known form, and the random error e may be assumed to follow normal distribution with zero mean and unknown variance σ^2. In this setup, the nonlinear least squares estimation of θ coincides with the

* Probal Chaudhuri is Member of the Scientific Staff, Division of Theoretical Statistics and Mathematics, Indian Statistical Institute, Calcutta 700035, India. Per A. Mykland is Assistant Professor, Department of Statistics, University of Chicago, IL 60637. Chaudhuri's research was supported in part by a Wisconsin Alumni Research Foundation Grant from University of Wisconsin. Mykland's research was supported in part by National Science Foundation Grant DMS 89-02667. The authors gratefully acknowledge many stimulating discussions with Douglas M. Bates, Edward W. Frees, and C. F. Jeff Wu. The authors also thank the editor, an associate editor, and several referees for helpful comments that led to important additions, deletions, and alterations in various drafts of the article.

maximum likelihood estimation, and the unknown standard deviation σ can be treated as a nuisance parameter. We are implicitly assuming that the distribution of e does not depend on X, so that X influences Y only through its mean.

Example 1.3. In binary response experiments, Y takes only two values—0 and 1—and one chooses an appropriate model for the conditional probability function $p(x) = P(Y = 1 | X = x)$. In the standard linear logistic regression, logit $\{p(x)\} = \log\{p(x)(1 - p(x))^{-1}\}$ is assumed to be linear function of x, whereas in the linear probit model, probit $\{p(x)\} = \Phi^{-1}\{p(x)\}$, is assumed to be a linear function of x, with Φ^{-1} the inverse of the standard normal cumulative distribution function. Numerous papers on bioassay have considered binary response experiments involving logistic and probit models; some very illustrative examples were mentioned and studied in Cox and Snell (1989) and McCullagh and Nelder (1989).

Example 1.4. McCullagh and Nelder (1989, chap. 8) considered some interesting biological experiments in which the response Y follows gamma distribution with density of the form $\{\Gamma(\nu)\}^{-1}(\alpha^{-1}\nu)^{\nu}y^{\nu-1}\exp(-\alpha^{-1}\nu y)$ such that y, α $\nu > 0$. The value of the parameter α depends on the value of X chosen by the experimenter, and the nature of this dependence is modeled using suitable functions.

Examples 1.3 and 1.4 are special cases of a very broad class of models popularly known as generalized linear models (GLM's) (see McCullagh and Nelder 1989; Nelder and Wedderburn 1972), where the conditional distribution of Y given $X = x$ is assumed to be in an exponential family with $f(y|\theta, \phi, x) = \exp[\{a(\phi)\}^{-1}\{y\tau - b(\tau)\} + c(y, \phi)]$. Here a, b, and c are all scalar functions and τ is another real-valued function of x and θ with some appropriate linear structure. For instance, when we have the canonical link function, τ can be expressed as $\tau = \tau(\theta, x) = \langle\theta, \mathbf{B}(x)\rangle$, where $\langle\, ,\rangle$ denotes the usual Euclidean inner product, and if the parameter space $\Theta \subseteq R^d$, then $\mathbf{B}$ will be a known R^d-valued function on Ω. Designing optimal binary response experiments has been investigated by Abdelbasit and Plackett (1983), Khan and Yazdi (1988), Minkin (1987), and others. Sometimes the response in the experiment occurs in the form of counts that have no upper bound. The Poisson regression model, which is frequently used in such cases (see, for example, Behnken and Watts 1972), is also a member of the GLM class. Another regression-type model, which has drawn some attention from people working on nonlinear experiments and belongs to this class, is the inverse Gaussian regression model (see, for example, Fries and Bhattacharya 1986).

In this article we will assume that the parameterization in the model $f(y|\theta, \phi, x)$ is smooth and regular in the sense that $\log\{f(y|\theta, \phi, x)\}$ is differentiable in both θ and ϕ and the associated Fisher information matrix exists finitely. It will be appropriate to note at this point that often in practice (in particular, in all of the preceding examples and families of models), the nuisance parameter ϕ has no influence on the computation of the maximum likelihood estimate for θ, and the Fisher information matrix computed from the complete model has a block diagonal form with two blocks corresponding to the efficient scores associated with θ and ϕ. The efficient scores are obtained by differentiating $\log\{f(y|\theta, \phi, x)\}$ with respect to θ and ϕ, and this block diagonal structure of the Fisher information matrix is a consequence of the orthogonality of the two scores associated with the parameter of principal interest and the nuisance parameter. As a matter of fact, for standard nonlinear regression problems and in the case of GLM's, the only contribution of the nuisance parameter to the score associated with θ is in the form of a nonrandom scalar multiple. From now on, in view of these simple and well-known yet quite crucial facts, we will ignore the presence of the parameter ϕ and write our model as $f(y|\theta, x)$. This is primarily to keep our notations simple as we concentrate on the efficient designing of experiments and estimation of θ using the maximum likelihood technique. We will assume that the parameter space Θ is an open convex subset of R^d and write $\mathbf{I}(\theta, x)$ to denote the Fisher information matrix, which is formally defined as $\mathbf{I}(\theta, x) = \int_{\mathcal{R}}[\nabla \log\{f(y|\theta, x)\}][\nabla \log\{f(y|\theta, x)\}]^T f(y|\theta, x)\mu(dy)$. Here ∇ is the usual gradient operator corresponding to differentiation with respect to θ and μ is the standard counting measure or the standard Lebesgue measure depending on whether the response space $\mathcal{R}$ is a countable set (finite or infinite) or an interval on the real line. All vectors in this article are column vectors unless specified otherwise, and the superscript T is used to indicate the transpose of vectors and matrices.

Following Kiefer (1959; 1961a,b), Kiefer and Wolfowitz (1959), and others, we define the design space $\mathcal{D}(\Omega)$ as the collection of all probability measures on Ω. According to the criterion of D-optimality, which has been used by several authors starting from Wald (1943) in a variety of situations, $\xi^* \in \mathcal{D}(\Omega)$ is a "locally optimal design" (see Chernoff 1953) at θ if $\det\{\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)\} = \sup_{\xi\in\mathcal{D}(\Omega)} \det\{\int_\Omega \mathbf{I}(\theta, x)\xi(dx)\}$, where "det" stands for the determinant of a matrix. Clearly, such a ξ^* will always exist in an application, as Ω will typically be a finite set or some nice subset of an Euclidean space and $\mathbf{I}(\theta, x)$ will be continuous in x for standard models. But for a nonlinear experiment with $\mathbf{I}(\theta, x)$ a nonlinear function of θ, ξ^* will depend on the unknown parameter θ. Cochran (1973, pp. 771–772) described this dependency: "You tell me the value of θ, and I promise to design the best experiment for estimating θ" (see also Myers, Khuri, and Carter 1989, pp. 143–144). An interesting technical study of the Fisher information in nonlinear regression can be found in Pazman (1989).

Box and Lucas (1959) considered some standard nonlinear regression models and tried to choose the design points $X_1, X_2, \ldots, X_n$ by maximizing the determinant of the total Fisher information $\sum_{i=1}^n \mathbf{I}(\theta, X_i)$ at some initial estimate $\theta = \theta_0$. Abdelbasit and Plackett (1983) and Minkin (1987) took a similar strategy in designing experiments with binary response following linear logistic regression models. An obvious practical drawback of this approach, noted by several authors, is that the prior estimate θ_0 may be far from true θ and the behavior of the locally optimal design may be quite sensitive to even small perturbations in the parameter value. One way to remedy or at least alleviate this problem is to adopt a multistage sequential design (see, for example, Abthat will ensure desirable behavior of the maximum likelihood estimate and convergence of the design to an optimal one. These conditions are easy to verify, and they will be satisfied for standard models used in practice.

3. ASYMPTOTIC OPTIMALITY OF THE DESIGN AND THE BEHAVIOR OF THE MAXIMUM LIKELIHOOD ESTIMATE

We first focus our attention on the asymptotic optimality of the chosen design. Obviously, the performance of the design will depend on our choice of n_1, the initial design points $X_1, X_2, \ldots, X_{n_1}$, and the estimates θ_i^*'s. Sufficient conditions that ensure the convergence of the chosen design to a D-optimal one as n grows to infinity follow.

Condition 3.1 (choice of the initial design). n_1 tends to infinity as n tends to infinity. Further, the initial design points $X_1, X_2, \ldots, X_{n_1}$ are chosen in such a way that the smallest eigenvalue of the matrix $n_1^{-1} \sum_{i=1}^{n_1} \mathbf{I}(\theta, X_i)$ remains bounded away from 0 as n tends to infinity for any $\theta \in \Theta$.

Condition 3.2 (the relative size of the initial experiment). The fraction n_1/n tends to 0 as n tends to infinity.

Condition 3.3 (a consistency condition). For any $\varepsilon > 0$, $\max_{n_1 \le i < n} P_\theta(|\theta_i^* - \theta| > \varepsilon)$ tends to 0 as n tends to infinity.

Condition 3.4 (a stability condition). For $n_1 < k < n$, let U_k denote the product of the determinants $\prod_{i=n_1+1}^{k} \det\{\sum_{r=1}^{i} \mathbf{I}(\theta_{i-1}^*, X_r)\} \det\{\sum_{r=1}^{i} \mathbf{I}(\theta_i^*, X_r)\}^{-1}$. Then, for any $\varepsilon > 0$, $\max_{n_1 < k < n} P_\theta(U_k > 1 + \varepsilon)$ tends to 0 as n tends to infinity.

Theorem 3.5. Assume Conditions 2.1, 2.2, and 3.1 through 3.4. If design points are chosen following our scheme at the sequential stage of the experiment, then $n^{-1} \sum_{i=1}^{n} \mathbf{I}(\theta, \mathbf{X}_i)$ will converge to $\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$ *in probability* as n tends to infinity. Here ξ^* is a locally D-optimal design at θ, as described in Section 1.

We next turn our attention to the behavior of the maximum likelihood estimate $\hat{\theta}_n$. At this point we need to introduce some conditions on the model $f(y|\theta, x)$. Recall from Section 1 that the parameter space Θ is assumed to be an open convex subset of R^d. From now on we will write $|\cdot|$ to denote the usual Euclidean norm of vectors and matrices.

Condition 3.6. The support of $f(y|\theta, x)$ does not depend on θ or x. Further, for every fixed $x \in \Omega$ and $y \in \mathcal{R}$, $\log\{f(y|\theta, x)\}$ is thrice continuously differentiable in θ.

Condition 3.7. Let $\nabla \log\{f(y|\theta, x)\} = \mathbf{G}(y, \theta, x)$ be the gradient vector obtained by computing the first-order partial derivatives of $\log\{f(y|\theta, x)\}$ with respect to θ. Then $\mathbf{G}(y, \theta, x)$ satisfies $\int_{\mathcal{R}} \mathbf{G}(y, \theta, x) f(y|\theta, x)\mu(dy) = 0$ and $\sup_{x\in\Omega} \int_{\mathcal{R}} |\mathbf{G}(y, \theta, x)|^{2+t} f(y|\theta, x)\mu(dy) < \infty$ for some $t > 0$.

Condition 3.8. Let $\mathbf{H}(y, \theta, x)$ denote the $d \times d$ Hessian matrix of $\log\{f(y|\theta, x)\}$ obtained by computing the second-order partial derivatives with respect to θ. Then $\mathbf{H}(y, \theta, x)$ satisfies $\int_{\mathcal{R}} \mathbf{H}(y, \theta, x) f(y|\theta, x)\mu(dy) = -\int_{\mathcal{R}} \{\mathbf{G}(y, \theta, x)\}\{\mathbf{G}(y, \theta, x)\}^T f(y|\theta, x)\mu(dy) = -\mathbf{I}(\theta, x)$, and $\sup_{x\in\Omega} \int_{\mathcal{R}} |\mathbf{H}(y, \theta, x)|^2 f(y|\theta, x)\mu(dy) < \infty$.

Condition 3.9. For every $\theta \in \Theta$, there is an open neighborhood $N(\theta)$ of θ and a nonnegative random variable $K(y, \theta, x)$ such that $\sup_{x\in\Omega} \int_{\mathcal{R}} K(y, \theta, x) f(y|\theta, x)\mu(dy) < \infty$, and each of the third-order partial derivatives of $\log\{f(y|\theta', x)\}$ with respect to θ' is dominated by $K(y, \theta, x)$ for all $\theta' \in N(\theta)$.

Theorem 3.10. Assume that in addition to conditions assumed in Theorem 3.5, Conditions 3.6 through 3.9 hold. Then there is a consistent choice of the maximum likelihood estimate $\hat{\theta}_n$ of θ such that, as n tends to infinity, the distribution of $n^{1/2}(\hat{\theta}_n - \theta)$ converges *weakly* to a d-dimensional normal distribution with zero mean and $\{\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)\}^{-1}$ as the variance–covariance matrix.

Corollary 3.11. Suppose that all the conditions assumed in Theorems 3.5 and 3.10 hold and let $\hat{\theta}_n$ be a consistent choice of the maximum likelihood estimate. Then, as n grows to infinity, the estimated average Fisher information $n^{-1} \sum_{i=1}^{n} \mathbf{I}(\hat{\theta}_n, X_i)$ converges *in probability* to the D-optimal Fisher information $\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$. Further, the asymptotic distribution of $\{\sum_{i=1}^{n} \mathbf{I}(\hat{\theta}_n, X_i)\}^{1/2}(\hat{\theta}_n - \theta)$ is d-variate normal with zero mean and the $d \times d$ identity matrix as the variance–covariance matrix.

Theorem 3.5 provides an answer to question 2 raised in the preceding section by asserting the convergence of our chosen design to a D-optimal one as the number of trials grows to infinity. Questions 1 and 3 are answered in the affirmative by Theorem 3.10 and Corollary 3.11. Theorem 3.10 ensures the existence of at least one choice of the maximum likelihood estimate $\hat{\theta}_n$ that is $n^{1/2}$-consistent, asymptotically normal, and first-order efficient. In some situations the consistent choice of the maximum likelihood estimate may very well be a local, rather than a global, maximizer of the likelihood. However, if we consider a model in the GLM family with the canonical link, the strict concavity of the log-likelihood function implies the uniqueness of the root of the likelihood equation and Theorem 3.10 guarantees the asymptotic optimality of the unique root. In many other problems involving models with special structures, one can work out appropriate conditions for the asymptotic uniqueness of the maximum likelihood estimate. For example, in the case of a nonlinear regression model $Y = g(\theta, X) + e$, appropriate regularity conditions on the regression function g will ensure optimal asymptotic properties of a global nonlinear least squares estimate of θ. In view of Corollary 3.11, we can construct confidence ellipsoids for θ using $\hat{\theta}_n$ and $\sum_{i=1}^{n} \mathbf{I}(\hat{\theta}_n, X_i)$. Any such confidence ellipsoid will asymptotically have the right coverage probability, and the convergence of our design to a D-optimal one implies that the asymptotic volume of the ellipsoid will be first-order optimal. When there is an unknown nuisance parameter (e.g., σ in the case of usual nonlinear regression and ϕ in the case of GLM), we will have to estimate it from the data to construct confidence sets for θ. Maximum likelihood techniques can be used for this purpose as well.

The choice of initial design points satisfying Condition 3.1 is quite simple for models typically encountered in practice. For a typical member in the GLM class with the canonical link, Condition 3.1 is equivalent to the condition that the smallest eigenvalue of the matrix $n_1^{-1} \sum_{i=1}^{n_1} \{\mathbf{B}(X_i)\}\{\mathbf{B}(X_i)\}^T$ remains bounded away from 0 as n_1 tends to infinity irrespective of the value of the parameter θ as long as the conditional variance function $\text{var}_\theta(Y|X = x) = b''(\langle\theta, \mathbf{B}(x)\rangle)a(\phi)$ is positive (see McCullagh and Nelder 1989, p. 29). A similar assertion holds for the regression setup considered by Woodroofe (1989) and Wu (1985). In fact an evenly distributed choice of the X_i's over Ω will frequently be sufficient to ensure Condition 3.1. Note that the two conditions—namely, n_1 tends to infinity and the fraction n_1/n tends to 0—imply that the size of the initial experiment should be allowed to grow with the increase in the size of the entire experiment, but the sequential stage of the experiment must play the dominant role. In a large-scale experiment, most of the trials are to be conducted in an adaptive sequential fashion, whereas the initial static experiment should consist of only a small fraction of the total number of trials. The size of the initial experiment and the smallest eigenvalue of the matrix $n_1^{-1} \sum_{i=1}^{n_1} \mathbf{I}(\theta, X_i)$ will influence the statistical stability and the precision of the initial estimate $\theta^*_{n_1}$. This suggests that Condition 3.1 has a subtle connection with Conditions 3.2 and 3.4.

We will now prove the existence of estimates θ_i^*'s satisfying Conditions 3.3 and 3.4 by means of explicit construction. Let $1 < n_1 < n_2 < n$ be such that n_1 tends to infinity and n_1/n_2 remains bounded away from 0 as n tends to infinity. Then, in view of Condition 3.1, the smallest eigenvalue of $n_2^{-1} \sum_{i=1}^{n_2} \mathbf{I}(\theta, X_i)$ must remain bounded away from 0 as n tends to infinity. For $n_1 \le i \le n_2$ let us define $\theta_i^* = \hat{\theta}_i$, where $\hat{\theta}_i$ is the maximum likelihood estimate of θ based on $(Y_1, X_1), (Y_2, X_2), \ldots, (Y_i, X_i)$, as mentioned in Section 2. When $n_2 < i < n$, we set $\theta_i^* = \hat{\theta}_{n_2} = \theta^*_{n_2}$. Using arguments virtually identical to the proof of the consistency of $\hat{\theta}_n$ (see the proof of Theorem 3.10 in the Appendix), it is straightforward to verify that the consistency Condition 3.3 will hold for some judicious choice of the maximum likelihood estimate whenever Condition 3.1 and Conditions 3.6 through 3.9 hold. On the other hand, in view of Condition 2.2 $\mathbf{I}(\theta, x)$ will be uniformly continuous in its two arguments if θ varies in a compact set contained in Θ and x runs over the compact space Ω. In particular this implies that whenever Condition 3.1 holds and $n_1 \le i < n$, $\sum_{r=1}^{i} \mathbf{I}(\theta_i^*, X_r)$ will be positive definite, and for $n_1 \le i \le n_2$ the smallest eigenvalue of $i^{-1} \sum_{r=1}^{i} \mathbf{I}(\theta_i^*, X_r)$ will remain bounded away from 0 provided that θ_i^* falls within a suitable neighborhood of true θ. Note at this point that Condition 3.4 will be automatically satisfied if the model is locally linear (i.e., in a neighborhood of the true parameter, the Fisher information is constant with respect to the value of θ and depends only on x) and the estimates θ_i^*'s are contained in a small neighborhood of true θ, so that U_k's become identically equal to 1 for $n_1 < k < n_2$. Consider now a collection of estimates θ_i^*'s that satisfy Condition 3.3 and $\theta_i^* = \theta^*_{n_2}$ for $n_2 \le i < n$. These estimates may or may not be maximum likelihood estimates. We define E_n as the set of all indices i such that the consecutive estimates θ^*_{i-1} and θ_i^* are distinct. Formally, we can write $E_n = \{i \mid n_1 < i \le n_2 \text{ and } \theta^*_{i-1} \ne \theta^*_i\}$. If m is a fixed positive integer that does not depend on n, then it is trivial to check that the stability Condition 3.4 will be satisfied whenever $|E_n| \le m$ for all large n (e.g., we can force n_2 to be smaller than or equal to $n_1 + m$). In view of this construction, Conditions 3.3 and 3.4 now appear to be practical advantages instead of technical barriers. It is quite clear that these conditions will be satisfied if we do not update the estimate θ_i^* too frequently at the sequential stage of the design in cases where n (and hence n_1) is large. This will save lots of computations in the actual implementation of the scheme for large n. Besides, the flexible nature of the scheme allows us to work with θ_i^*'s that may be less efficient than the maximum likelihood estimate but easier to compute and hence easier to update. Any consistent and suitably stable choice of θ_i^*'s will ensure the convergence of the design to an optimal one. But the impact of the choice of n_1 and θ_i^*'s on the rate of this convergence and how close we can get to the optimal design for finite n are issues yet to be investigated.

Conditions 3.6 through 3.9 are standard Cramer-type conditions on the model that have been used by several authors in deriving the asymptotic properties of maximum likelihood estimates. For a typical model in the GLM class, these conditions are satisfied. For a nonlinear regression problem involving normal error, these conditions translate into some regularity conditions on the regression function g ensuring the consistency and the asymptotic normality of the nonlinear least squares estimate (see, for example, Gallant 1987, Jenrich 1969, Seber and Wild 1989, and Wu 1981).

4. CONCLUDING REMARKS

As noted previously, Wu (1985) and later Woodroofe (1989) used some asymptotic results on sums of martingale difference sequences developed in Lai and Wei (1982) (see also Wei 1985) in the context of a sequential design for a regression problem. In the proofs of our theorems, we use Burkholder's inequality (Burkholder 1973) and some results from martingale limit theory given in Hall and Heyde (1980). The key observation, as discussed in Section 2, is that despite the dependent nature of the data arising from a sequentially designed experiment, the likelihood remains in the product form. As a result, various derivatives of the logarithm of the likelihood with some suitable adjustments will give rise to sums of martingale difference sequences, which form the rows of certain triangular arrays. The asymptotic behavior of maximum likelihood estimates in dependent processes has been extensively studied in the literature (see, for example, Basawa and Prakasa Rao 1980; Prakasa Rao 1987; and Sweeting 1980, 1983).

Ford et al. (1989, p. 51) remarked that the relevance of asymptotics is less clear in the case of a sequential experiment than it is for a static experiment and explicitly mentioned that the only asymptotic that one might contemplate for a sequential experiment is the evolution of the design as the number of trials grows. Further, they emphasized that even if ξ^* is known, it is not clear that the matrix $\{n \int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)\}^{-1}$ is useful as an approximate variance–covariance matrix of the maximum likelihood estimate $\hat{\theta}_n$ for finite n. Theorems 3.5 and 3.10 and Corollary 3.11 are products of a thorough investigation into the issues raised by them. In technical terms, the convergence of the random average Fisher information $n^{-1}\sum_{i=1}^{n}\mathbf{I}(\theta, X_i)$ into the deterministic positive definite matrix $\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$ is an ergodicity condition that plays a fundamental role in establishing the asymptotic normality and the first-order optimality of $\hat{\theta}_n$ via martingale central limit theorem. From a practical standpoint, these asymptotic results validate statistical inference based on $\hat{\theta}_n$, ensure the optimality of the design, and guarantee a high degree of precision in parameter estimates and confidence sets.

One of several nice features of the criterion of D-optimality based on the Fisher information matrix is that it leads to designs that are invariant under smooth and regular reparameterization of models. But the effect of the curvature of a nonlinear model on parameter estimates is an important issue in finite sample situations. Bates and Watts (1980, 1981), Hamilton and Watts (1985), Hamilton, Watts, and Bates (1982), Seber and Wild (1989), and others have explored this problem in the context of standard nonlinear regression. It will be interesting to investigate sequential designs that allow for curvature effects, and one may hope to achieve a second-order asymptotic optimality there.

APPENDIX: PROOFS

We begin by making some elementary but critical observations and proving some preliminary results.

Lemma A.1. The function $h(\mathbf{A}) = -\log\{\det(\mathbf{A})\}$, where $\mathbf{A}$ is a symmetric $d \times d$ positive definite matrix, is a strictly convex function. In other words, for $0 < \alpha < 1$ and two positive definite matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{A} \neq \mathbf{B}$, we have $h\{\alpha\mathbf{A} + (1-\alpha)\mathbf{B}\} < \alpha h(\mathbf{A}) + (1-\alpha)h(\mathbf{B})$.

Proof. The assertion is trivially true when both $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices. Also, it is a well-known fact in matrix algebra that for every pair of positive definite real symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, there is a nonsingular matrix $\mathbf{C}$ such that $\mathbf{A} = \mathbf{C}^T\mathbf{D}\mathbf{C}$ and $\mathbf{B} = \mathbf{C}^T\mathbf{C}$, where $\mathbf{D}$ is a diagonal matrix (see, for example, Rao 1973, p. 41). The lemma is immediate after using this fact.

This lemma ensures that as long as there is a $\xi \in \mathcal{D}(\Omega)$ such that $\int_\Omega \mathbf{I}(\theta, x)\xi(dx)$ is nonsingular, the D-optimal Fisher information matrix $\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$ is uniquely defined, even though the D-optimal design ξ^* on Ω may not be unique. In particular, this makes the assertions made in Theorems 3.5 and 3.10 meaningful. Further, we have the following lemma.

Lemma A.2. Let $\{\theta_n\}$ be a sequence of points in Θ such that, as n tends to infinity, θ_n converges to $\theta \in \Theta$. Let ξ_n^* be a locally D-optimal design associated with θ_n and let ξ^* be that associated with θ. Then, under Conditions 2.1 and 2.2, the matrix $\int_\Omega \mathbf{I}(\theta_n, x)\xi_n^*(dx)$ converges to $\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$ as n grows to infinity, provided that $\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$ is nonsingular.

Proof. It is straightforward and easy to see using the weak compactness of probability measures defined on a compact metric space and the continuity of $\mathbf{I}(\theta, x)$ that $\det\{\int_\Omega \mathbf{I}(\theta_n, x)\xi_n^*(dx)\}$ will converge to $\det\{\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)\}$ as n tends to infinity. Further, under Conditions 2.1 and 2.2, any subsequence of the sequence $\{\int_\Omega \mathbf{I}(\theta_n, x)\xi_n^*(dx)\}$ will have a further subsequence that will be convergent with the limit $\int_\Omega \mathbf{I}(\theta, x)\xi(dx)$ for some $\xi \in D(\Omega)$. The proof of the lemma is now complete using Lemma A.1, which ensures that such a ξ must be a locally D-optimal design corresponding to θ.

We now state a fact, which is a consequence of the well-known inequality $\log(1+x) < x$ that holds for all $x > 0$.

Fact 5.3. Let $1 < n_1 < n$ be integers (n_1 may be a function of n) such that n_1/n tends to 0 as n tends to infinity. Then, as n grows to infinity, the sum $\sum_{i=n_1}^{n} i^{-1}$ diverges to infinity.

Proof of Theorem 3.5. For $n_1 \le i < n$, let ξ_i^* denote a locally D-optimal design corresponding to θ_i^*. We will write $\mathbf{I}_i(\theta')$ for the matrix $i^{-1}\sum_{r=1}^{i}\mathbf{I}(\theta', X_r)$, where $\theta' \in \Theta$, and $\mathbf{I}_i^*$ will denote the matrix $\int_\Omega \mathbf{I}(\theta_i^*, x)\xi_i^*(dx)$. Fix an η such that $0 < \eta < \frac{1}{4}$ and define S_n to be the collection of all positive integers i such that $n_1 \le i < n$ and $\det\{\mathbf{I}_{i+1}(\theta_i^*)\} \ge (1-\eta)\det(\mathbf{I}_i^*)$. Then it follows from inequality 2.3.12 in Theorem 2.3.4 in Kiefer (1961b, p. 389), via arguments in the proof of Theorem 1 in Wynn (1970, pp. 1658–1660), that $\det\{\mathbf{I}_{i+1}(\theta_i^*)\}\det\{\mathbf{I}_i(\theta_i^*)\}^{-1} \ge 1 + i^{-1}\rho$ for any $i \notin S_n$ and $n_1 \le i < n$ provided that n and n_1 are suitably large and $\det\{\mathbf{I}_i(\theta_i^*)\} > 0$ (see inequalities 3.5, 3.7, and so on in Wynn 1970, pp. 1658–1659). Here ρ is a positive constant that depends on η. Now, Condition 3.2 and Fact 5.3 imply that the product $\prod_{i=n_1}^{n-1}(1 + i^{-1}\rho)$ must diverge to infinity as n tends to infinity for any $\rho > 0$. Hence in view of Conditions 2.1, 2.2, and 3.1 through 3.4, we can conclude that $\lim_{n\to\infty} P_\theta\,(S_n \textit{ is empty}) = 0$. At this point, Lemma 5.2, Conditions 3.3 and 3.4, and some minor modification of arguments starting from inequality 3.8 in the proof of Theorem 1 in Wynn (1970, pp. 1659–1660) using some straightforward algebra imply that $\lim_{n\to\infty} P_\theta[\det\{\mathbf{I}_n(\theta_{n-1}^*)\} > (1-4\eta)\det(\mathbf{I}^*)] = 0$, where $\mathbf{I}^*$ is the D-optimal Fisher information matrix $\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$. In other words, $\det\{\mathbf{I}_n(\theta_{n-1}^*)\}$ converges to $\det(\mathbf{I}^*)$ in probability as n tends to infinity. The consistency Condition 3.3, the compactness of Ω, and the continuity of $\mathbf{I}(\theta, x)$ now yield the weak convergence of $\det\{\mathbf{I}_n(\theta)\}$ into $\det(\mathbf{I}^*)$. Finally, the proof of the theorem is complete, exploiting the observation made in Lemma A.1 and the arguments used in the proof of Lemma A.2.

Proof of Theorem 3.10. We begin by proving that there exists a maximizer $\hat{\theta}_n$ (possibly a local one) of the likelihood that is consistent for θ. In view of Condition 3.7 and the product form of the likelihood, the gradient $\sum_{i=1}^{n}\mathbf{G}(Y_i, \theta, X_i)$ of the log-likelihood is a sum of square integrable martingale difference sequence if we introduce the increasing sequence of σ-fields $\mathcal{F}_{ni}$'s, where $1 \le i \le n$ and $\mathcal{F}_{ni}$ is generated by $Y_1, Y_2, \ldots, Y_i$. Here, as n and i vary, $\mathbf{G}(Y_i, \theta, X_i)$'s generate a triangular array. Using Burkholder's inequality (Burkholder 1973, Hall and Heyde 1980), we conclude that $\sum_{i=1}^{n}\mathbf{G}(Y_i, \theta, X_i)$ is $O(n^{1/2})$ in probability as n tends to infinity. Next, note that if we compute the Hessian matrix of the log-likelihood and consider $\sum_{i=1}^{n}\{\mathbf{H}(Y_i, \theta, X_i) + \mathbf{I}(\theta, X_i)\}$, then Condition 3.8 implies that it is another sum of square integrable martingale difference sequence, which is also $O(n^{1/2})$ in probability. Theorem 3.5 now ensures the weak convergence of $n^{-1}\sum_{i=1}^{n}\mathbf{H}(Y_i, \theta, X_i)$ into $-\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$ as n grows to infinity. Because $\int_\Omega \mathbf{I}(\theta, x) \times \xi^*(dx)$ must be a positive definite matrix in view of Condition 3.1, a third-order Taylor expansion of the log-likelihood and Condition 3.9 ensure that the log-likelihood will be a locally concave function of the parameter in an appropriately small neighborhood of true θ with probability tending to 1 as n tends to infinity. In fact we can choose a $\delta_0 > 0$ such that as n tends to infinity, the log-likelihood will be concave in a neighborhood of radius δ_0 around true θ, and for any δ satisfying $0 < \delta \le \delta_0$ the likelihood equation will have a root (i.e., a local maximizer of the log-likelihood) in the neighborhood of true θ having radius δ with probability tending to 1.

We next establish the asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta)$, where $\hat{\theta}_n$ is a consistent choice of the maximum likelihood estimate. From Corollary 3.1 in Hall and Heyde (1980, p. 58) and the Cramer–Wold device (see, for example, Billingsley 1968, pp. 48–49), it follows that $n^{-1/2} \sum_{i=1}^{n} \mathbf{G}(Y_i, \theta, X_i)$ converges weakly to a d-dimensional normal random vector with zero mean and $\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$ as the variance–covariance matrix. This is because of our Theorem 3.5 (verifying the condition on the conditional variance–covariance process in Hall and Heyde 1980, p. 59) and Condition 3.7 (verifying the conditional Lindeberg condition in Hall and Heyde 1980, p. 58). A second-order Taylor expansion of the gradient of the log-likelihood using Conditions 3.8 and 3.9 and the weak convergence of $n^{-1} \sum_{i=1}^{n} \mathbf{H}(Y_i, \theta, X_i)$ into the negative definite matrix $-\int_\Omega \mathbf{I}(\theta, x)\xi^*(dx)$ now yields the assertion in the theorem. The argument here is quite similar to the proof of consistency in the previous paragraph. First the $n^{1/2}$-consistency of $\hat{\theta}_n$ is established; the limit law then follows.

[*Received October 1990. Revised September 1992.*]

REFERENCES

Abdelbasit, K. M., and Plackett, R. L. (1983), "Experimental Design for Binary Data," *Journal of the American Statistical Association*, 78, 90–98.

Atkinson, A. C. (1982), "Developments in the Design of Experiments," *International Statistical Review*, 50, 161–177.

Basawa, I. V., and Prakasa Rao, B. L. S. (1980), *Statistical Inference for Stochastic Processes*, London: Academic Press.

Bates, D. M. (1983), "The Derivative of | X'X | and Its Uses," *Technometrics*, 25, 373–376.

Bates, D. M., and Watts, D. G. (1980), "Relative Curvature Measures of Nonlinearity" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 42, 1–25.

——— (1981), "Parameter Transformations for Improved Approximate Confidence Regions in Nonlinear Least Squares," *The Annals of Statistics*, 9, 1152–1167.

——— (1988), *Nonlinear Regression, Analysis and Its Applications*, New York: John Wiley.

Behnken, D. W., and Watts, D. G. (1972), "Bayesian Estimation and Design of Experiments for Growth Rates When Sampling From the Poisson Distribution," *Biometrics*, 28, 999–1009.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Borkar, V., and Varaiya, P. (1979), "Adaptive Control of Markov Chains I: Finite Parameter Set," *IEEE Transactions on Automatic Control*, AC-24, 953–958.

——— (1982), "Identification and Adaptive Control of Markov Chains," *SIAM Journal on Control and Optimization*, 20, 470–489.

Box, G. E. P., and Hunter, W. G. (1965), "Sequential Design of Experiments for Nonlinear Models," in *Proceedings of the IBM Scientific Computing Symposium on Statistics, October 21-23, 1963*, pp. 113–137.

Box, G. E. P., and Lucas, H. L. (1959), "Design of Experiments in Nonlinear Situations," *Biometrika*, 49, 77–90.

Box, M. J. (1968), "The Occurrence of Replications in Optimal Designs of Experiments to Estimate Parameters in Nonlinear Models," *Journal of the Royal Statistical Society*, Ser. B, 30, 290–302.

——— (1970), "Some Experiences With a Nonlinear Experimental Design Criterion," *Technometrics*, 12, 569–589.

Burkholder, D. L. (1973), "Distribution Function Inequalities for Martingales," *The Annals of Probability*, 1, 19–42.

Carr, N. L. (1960), "Kinetics of Catalytic Isomerization of n-pentane," *Industrial and Engineering Chemistry*, 52, 391–396.

Chaloner, K. (1989), "Bayesian Design for Estimating the Turning Point of a Quadratic Regression," *Communications in Statistics, Part A—Theory and Methods*, 18, 1385–1390.

Chaloner, K., and Larntz, K. (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Planning and Inference*, 21, 191–208.

Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 30, 586–602.

——— (1975), "Approaches in Sequential Design of Experiments," in *A Survey of Statistical Design and Linear Models*, ed. J. N. Srivastava, New York: North-Holland, pp. 67–90.

Cochran, W. G. (1973), "Experiments for Nonlinear Functions" (R. A. Fisher Memorial Lecture), *Journal of the American Statistical Association*, 68, 771–781.

Cox, D. R., and Snell, E. J. (1989), *Analysis of Binary Data*, London: Chapman and Hall.

Draper, N. R., and Hunter, W. G. (1967a), "The Use of Prior Distributions in the Design of Experiments for Parameter Estimation in Nonlinear Situations," *Biometrika*, 54, 147–153.

——— (1967b), "The Use of Prior Distributions in the Design of Experiments for Parameter Estimation in Nonlinear Situations: Multiresponse case," *Biometrika*, 54, 662–665.

El-Fattah, Y. M. (1981a), "Recursive Algorithm for Adaptive Control of Finite Markov Chains," *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-11, 135–144.

——— (1981b), "Gradient Approach for Recursive Estimation and Control in Finite Markov Chains," *Advances in Applied Probability*, 13, 778–803.

Fedorov, V. V. (1972), *Theory of Optimal Experiments*, New York: Academic Press.

Ford, I. (1976), "Optimal Static and Sequential Design: A Critical Review," unpublished Ph.D. dissertation, University of Glasgow, Dept. of Statistics.

Ford, I., and Silvey, S. D. (1980), "A Sequentially Constructed Design for Estimating a Nonlinear Parametric Function," *Biometrika*, 67, 381–388.

Ford, I., Titterington, D. M., and Kitsos, C. P. (1989), "Recent Advances in Nonlinear Experimental Design," *Technometrics*, 31, 49–60.

Ford, I., Titterington, D. M., and Wu, C. F. J. (1985), "Inference and Sequential Design," *Biometrika*, 72, 545–551.

Fries, A., and Bhattacharya, G. K. (1986), "Optimal Design for an Inverse Gaussian Regression Model," *Statistics and Probability Letters*, 4, 291–294.

Gallant, A. R. (1987), *Nonlinear Statistical Models*, New York: John Wiley.

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, New York: Academic Press.

Hamilton, D. C., and Watts, D. G. (1985), "A Quadratic Design Criterion for Precise Estimation in Nonlinear Regression Models," *Technometrics*, 27, 241–250.

Hamilton, D. C., Watts, D. G., and Bates, D. M. (1982), "Accounting for Intrinsic Nonlinearity in Nonlinear Regression Parameter Inference Regions," *The Annals of Statistics*, 10, 386–393.

Hill, P. D. H. (1980), "*D*-Optimal Designs for Partially Nonlinear Regression Models," *Technometrics*, 22, 275–276.

Hill, W. J., and Hunter, W. G. (1974), "Design of Experiments for Subsets of Parameters," *Technometrics*, 16, 425–434.

Hohmann, G., and Jung, W. (1975), "On Sequential and Nonsequential *D*-Optimal Experiment Design," *Biometrisch Zeitschrift*, 17, 329–336.

Jenrich, R. J. (1969), "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

Khan, M.K., and Yajdi, A. A. (1988), "On *D*-Optimal Designs for Binary Data," *Journal of Statistical Planning and Inference*, 18, 83–91.

Khuri, A. I. (1984), "A Note on *D*-Optimal Designs for Partially Nonlinear Regression Models," *Technometrics*, 26, 59–61.

Kiefer, J. (1959), "Optimum Experimental Design" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 21, 272–319.

——— (1961a), "Optimum Designs in Regression Problems, II," *Annals of Mathematical Statistics*, 32, 298–325.

——— (1961b), "Optimum Experimental Designs V, With Applications to Systematic and Rotatable Designs," in *Proceedings of the Fourth Berkeley Symposium* (Vol. 1), pp. 381–405.

Kiefer, J., and Wolfowitz, J. (1959), "Optimum Designs in Regression Problems," *Annals of Mathematical Statistics*, 30, 271–294.

Kitsos, C. P. (1989), "Fully Sequential Procedures in Nonlinear Design Problems," *Computational Statistics and Data Analysis*, 8, 13–19.

Kumar, P. R. (1985), "A Survey of Some Results in Stochastic Adaptive Control," *SIAM Journal on Control and Optimization*, 23, 329–380.

Kumar, P. R., and Becker, A. (1982), "A New Family of Optimal Adaptive Controller for Markov Chains," *IEEE Transactions on Automatic Control*, AC-27, 137–146.

Kumar, P. R., and Lin, W. (1982), "Optimal Adaptive Controller for Unknown Markov Chains," *IEEE Transactions on Automatic Control*, AC-27, 765–774.

Lai, T. L., and Wei, C. Z. (1982), "Least Squares Estimates in Stochastic Regression Models With Applications to Identification and Control of Dynamic Systems," *The Annals of Statistics*, 10, 154–166.

Lauter, E. (1974), "A Method of Designing Experiments for Nonlinear Models," *Mathematische Operationsforschung und Statistik*, 5, 697–708.

McCormick, W. P., Mallik, A. K., and Reeves, J. H. (1988), "Strong Consistency of the MLE for Sequential Design Problems," *Statistics and Probability Letters*, 6, 441–446.

McCullagh, P. (1981), Discussion of "Randomized Allocation of Treatments in Sequential Experiments," by J. A. Bather, *Journal of the Royal Statistical Society,* Ser. B, 43, 286–287.

McCullagh, P., and Nelder, J. (1989), *Generalized Linear Models,* London: Chapman and Hall.

Minkin, S. (1987), "Optimal Designs for Binary Data," *Journal of the American Statistical Association,* 82, 1098–1103.

Myers, R. H., Khuri, A. I., and Carter, W. H. (1989), "Response Surface Methodology: 1966–1988," *Technometrics,* 31, 137–157.

Nelder, J. A., and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society,* Ser. A, 135, 370–384.

Pazman, A. (1989), "On Information Matrices in Nonlinear Experimental Design," *Journal of Statistical Planning and Inference,* 21, 253–263.

Prakasa Rao, B. L. S. (1987), *Asymptotic Theory of Statistical Inference,* New York: John Wiley.

Rasch, D. (1990), "Optimum Experimental Design in Nonlinear Regression," *Communications in Statistics, Part A—Theory and Methods,* 19, 4789–4806.

Rao, C. R. (1973), *Linear Statistical Inference,* New York: John Wiley.

Robertazzi, T. G., and Schwartz, S. C. (1989), "An Accelerated Sequential Algorithm for Producing *D*-Optimal Designs," *SIAM Journal on Scientific and Statistical Computing,* 10, 341–358.

Sato, M., Abe, K., and Takeda, H. (1982), "Learning Control of Finite Markov Chains With Unknown Transition Probabilities," *IEEE Transactions on Automatic Control,* AC-27, 502–505.

Seber, G. A. F., and Wild, C. J. (1989), *Nonlinear Regression,* New York: John Wiley.

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

Sweeting, T. J. (1980), "Uniform Asymptotic Normality of the Maximum Likelihood Estimator," *The Annals of Statistics,* 8, 1375–1381.

——— (1983), "On Estimator Efficiency in Stochastic Processes," *Stochastic Processes and Their Applications,* 15, 93–98.

Wald, A. (1943), "On the Efficient Design of Statistical Investigation," *Annals of Mathematical Statistics,* 14, 134–140.

Wei, C. Z. (1985), "Asymptotic Properties of Least Squares Estimates in Stochastic Regression Models," *The Annals of Statistics,* 13, 1498–1508.

Woodroofe, M. (1989), "Very Weak Expansions for Sequentially Designed Experiments," *The Annals of Statistics,* 17, 1087–1102.

Wu, C. F. J. (1981), "Asymptotic Theory of Nonlinear Least Squares Estimation," *The Annals of Statistics,* 9, 501–513.

——— (1985), "Asymptotic Inference From Sequential Design in a Nonlinear Situation," *Biometrika,* 72, 553–558.

Wu, C. F. J., and Wynn, H. P. (1978), "The Convergence of General Step Length Algorithms for Regular Optimum Design Criteria," *The Annals of Statistics,* 6, 1273–1285.

Wynn, H. P. (1970), "The Sequential Generation of *D*-Optimum Experimental Design," *Annals of Mathematical Statistics,* 41, 1655–1664.

——— (1972), "Results in the Theory and Construction of *D*-Optimum Experimental Design," *Journal of the Royal Statistical Society,* Ser. B, 34, 133–147.

Zacks, S. (1977), "Problems and Approaches in Design of Experiments for Estimation and Testing in Nonlinear Models," in *Proceedings of the 4th International Symposium on Multivariate Analysis,* ed. P. R. Krishnaiah, Amsterdam: North-Holland, pp. 209–223.