Title

Asymptotic Bayes Criteria for Nonparametric Response Surface Design

Authors
Toby Mitchell
Jerome Sacks
Donald Ylvisaker

Publication Date
2011-10-24

The Annals of Statistics
1994, Vol. 22, No. 2, 634–651

ASYMPTOTIC BAYES CRITERIA FOR NONPARAMETRIC RESPONSE SURFACE DESIGN

BY TOBY MITCHELL,[1,2] JEROME SACKS[3] AND DONALD YLVISAKER[4]

Oak Ridge National Laboratory, National Institute of Statistical Sciences and University of California, Los Angeles

This paper deals with Bayesian design for response surface prediction when the prior may be finite or infinite dimensional, the design space arbitrary. In order that the resulting problems be manageable, we resort to asymptotic versions of D-, G- and A-optimality. Here the asymptotics stem from allowing the error variance to be large. The problems thus elicited have strong game-like characteristics. Examples of theoretical solutions are brought forward, especially when the priors are stationary processes on an interval, and we give numerical evidence that the asymptotics work well in the finite domain.

1. Introduction. This paper deals with designs for situations in which a response function, or "signal," on the set T is obscured by noise and where observations may be replicated at the various points ("sites") in T. Thus, we think of production experiments which can be carried out, and even replicated, under a variety of conditions, or of experiments which make one or more assays on a given geological core sample. The set of sites under consideration may have characteristics which vary from case to case. We will, in fact, carry on the discussion with a completely general T, but our examples are quite specific.

The goal is to predict the underlying signal. The treatment is Bayesian in that knowledge of the signal is represented by a random function (stochastic process). Traditional Bayesian design theory, where the prior for the signal is a random finite linear combination of known functions, falls under this kind of formulation; see, for example, Pilz (1983), Chaloner (1984) and Bandemer, Nather and Pilz (1987). Here we consider mainly infinite-dimensional random functions as priors. This leads to a kind of "nonparametric" surface fitting method, so designated to distinguish it from the usual response surface methods, which

Received April 1991; revised September 1993.

[1]Toby Mitchell died on June 5, 1993 after a long illness. The other authors are deeply saddened by the loss of a dear friend, a valued colleague and an outstanding scientist.

[2]Research supported by the Applied Mathematical Sciences Research Program, Office of Energy Research, U.S. Department of Energy, through Contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

[3]Research supported in part by NSF Grant DMS-90-01726, NSA Grant MDA 904-89-H-2011 and by AT&T Affiliates Program 90. Support for computing was provided by Cray Research and National Center for Supercomputing Applications at the University of Illinois, Urbana–Champaign.

[4]Research supported in part by NSF Grants DMS-89-02494 and DMS-92-13264.

AMS 1991 *subject classifications.* Primary 62K05; secondary 62J02.

Key words and phrases. Bayesian design, asymptotics, D-, G- and A-optimality, stationary processes.

rely on approximation by low-degree polynomials or by other finite-dimensional families. Specifically, we assume that the prior is the one which derives from taking the signal to have the form $\beta + Z$, where β is a normal random variable and Z is a Gaussian process independent of β. In this respect, our treatment is similar to that of O'Hagan (1978), Wahba (1978), Sacks and Ylvisaker (1985), Steinberg (1985) and Sacks, Welch, Mitchell and Wynn (1989). A clear distinction can be drawn with the last reference: it deals with computer experimentation in which there is no noise, while we deal here with actual experimentation and its attendant variable responses.

In Section 2 we give three criteria for selecting prediction designs. In analogy with classical terminology we designate them as the D, G and A criteria. Direct optimization of any of these criteria poses a formidable problem. Indeed, this investigation began with an arduous and direct attempt to answer a very simple but basic question: If Z is an infinite-dimensional process on the unit interval, does one ever call for replication? It could be argued, for example, that the taking of observations in close proximity would prove to be an adequate substitute. If not, one would imagine that replication would be called for when Z is "smooth" enough. Our results will suggest that the latter is the right answer.

To make progress, we allow the variance of β and the variance of the noise to go to infinity while the variance of the Z process is held fixed. Through this device, manageable surrogates are obtained for the original optimality problems. Our interest lies in the form the new problems take, with the insights they provide and with the level of difficulty they present. We view it as a bonus if designs obtained in this way stand up to scrutiny in the finite domain as well, and some numerics are advanced in support of this.

Early reliance on the simplicity afforded by asymptotics for problems in the general area can be seen in Sacks and Ylvisaker (1966) and in Bickel and Herzberg (1979), as two examples. More recently, a report by Lim, Sacks, Studden and Welch (1988) has rekindled the thought that hard problems might have easier (asymptotic) siblings—the paper of Johnson, Moore and Ylvisaker (1990) is a good case in point. The last work considers designs when there is no noise and then replication is not even an issue.

The criteria we use do not generally have the properties which obtain in classical or other settings. In particular, the D and G criteria are equivalent in the approximate theory of Kiefer and Wolfowitz (1960), while in Johnson, Moore and Ylvisaker (1990) they complement each other in a natural way. Here we find that they differ in general, but there are cases where they agree, typically when the prior process $\beta + Z$ concentrates on functions that are suitably rough.

In Section 2 we give a full description of the prediction criteria, while the asymptotics for them are developed in Section 3. With the problems recast in this way we give some solutions and examples in Section 4. Section 5 numerically compares, in some simple cases, exact designs computed for the asymptotic criteria with exact designs computed for the original nonasymptotic ones. The former are much easier to compute and, in these examples at least, are quite efficient over wide ranges of the parameters of the prior process.

This paper could be made a good deal longer by pursuing some of the issues raised but not fully investigated. Our concern, rather, has been to point out how certain asymptotics produce a class of manageable Bayesian design problems from hard ones.

2. Prediction criteria. The prior distribution for the signal is given by the random function $\beta+Z$, where β is a Gaussian (normal) random variable with mean μ_β and variance σ_β^2, and Z is a Gaussian stochastic process independent of β, with $E(Z_t) = \mu(t)$, $\mathrm{Var}(Z_t) = \sigma^2(t)$ and covariance function $R(s,t) = \mathrm{Cov}(Z_s, Z_t)$, $t \in T$, $s \in T$. The signal is observed with error, at n (not necessarily distinct) sites $t_1, \dots, t_n$ in T. The n errors $\varepsilon_1, \dots, \varepsilon_n$ are taken to be additive, independent of the signal and of one another, and Gaussian with mean 0 and variance σ_ε^2. Thus the prior distribution for the ith observation is given by the Gaussian random variable

$$Y_i = \beta + Z_{t_i} + \varepsilon_i. \tag{2.1}$$

The objective is prediction of the signal on T, and the design element enters as the choice of the set $\{t_i\}$. Three prediction criteria are considered at length—they are termed the D (for determinant), G (for global) and A (for average) criteria. To describe them, let Var denote variance and let GenVar be generalized variance, that is, the determinant of the covariance matrix. First let T be a finite set. A design is called *G-optimum* if it minimizes

$$\max_{t_0 \in T} \mathrm{Var}\big(\text{signal at } t_0 \mid \text{chosen observations}\big); \tag{2.2}$$

a design is called *A-optimum* for a given probability distribution π on T if it minimizes

$$E_\pi \mathrm{Var}\big(\text{signal at } \Pi \mid \text{chosen observations}\big), \tag{2.3}$$

where Π has distribution π; a design is called *D-optimum* if it minimizes

$$\mathrm{GenVar}\big(\text{signal variables} \mid \text{chosen observations}\big). \tag{2.4}$$

Should T be an infinite set, "max" should be replaced by "sup" in (2.2) while for (2.3) one needs some structure to consider general distributions. As presently stated, (2.4) relies on the fact that there are finitely many signal variables. When this is so, one can replace (2.4) by the problem of maximizing

$$\mathrm{GenVar}(\text{chosen observations}). \tag{2.5}$$

(See Remark 3 about the equivalence of the two problems.) On the other hand, if T is infinite and a design is optimum under (2.5), it follows that the same design is optimum under (2.4) for any finite (and arbitrarily dense) subset of T that includes the design sites. It is easier to attack (2.5), but we prefer to think of the D criterion as it is stated in (2.4).

All three criteria (2.2)–(2.4) depend on the posterior covariance of the signal variables, which for Gaussian processes does not depend on the prior means.

Without loss of generality, then, we take $\mu_\beta = 0$ and $\mu(t) = 0$. For simplicity, we take $\sigma^2(t)$ to be constant, so $\mathrm{Cov}(Z_s, Z_t) = \sigma_Z^2 \rho(s,t)$, where ρ is a completely specified correlation function and σ_Z^2 is the (constant) variance of Z (but see Remark 4 in this connection). Letting $\sigma_\beta^2 = v\sigma_Z^2$ and $\sigma_\varepsilon^2 = \sigma^2\sigma_Z^2$, we see that σ_Z^2 appears in the relevant covariance matrices only as a constant multiplier. We therefore take $\sigma_Z^2 = 1$ without loss of generality, keeping in mind that v and σ^2 are the variances of β and ε, respectively, *relative* to the variance of Z.

To this point, problems (2.2), (2.3) and (2.5) are quite difficult and solutions depend heavily on v, σ^2 and the correlation function ρ. On the other hand, suitable asymptotics will produce much more tractable versions, and we turn to the mechanics of these now. We will let $v \to \infty$ and $\sigma^2 \to \infty$ while $\gamma = v/\sigma^2$ is held fixed. It will turn out that if $0 < \gamma < \infty$, there is a unique D-optimality problem in the limit. On the other hand, the G-optimality and A-optimality problems which result do depend on γ. It is of interest to note that if we take $v = 0$, so that the model is $Y = Z +$ noise, we get different problems but, at least mathematically, nothing new surfaces (Remark 2).

3. Asymptotics. We begin by observing that (2.1), along with the usual formula for conditional variance in the multivariate normal, allows us to write the variance in (2.2) as

$$v^*(t_0) = (v+1) - \left(v1_n' + \rho_0'\right)\left(vJ_n + \rho + \sigma^2 I_n\right)^{-1}(v1_n + \rho_0), \tag{3.1}$$

where ρ_0 is the n-vector of correlations between Z_{t_0} and $Z_{t_i}, i = 1, 2, \ldots, n, \rho$ is the correlation matrix of $Z_{t_1}, \ldots, Z_{t_n}$, 1_n is the n-vector of 1's and $J_n = 1_n 1_n'$. Now let $\gamma = v/\sigma^2$ and $\alpha = 1/\sigma^2$. Then

$$v^*(t_0) = \alpha^{-1}\left[(\gamma+\alpha) - (\gamma 1_n' + \alpha\rho_0')(\gamma J_n + \alpha\rho + I_n)^{-1}(\gamma 1_n + \alpha\rho_0)\right]. \tag{3.2}$$

if $W = I_n + \gamma J_n$, then

$$(W + \alpha\rho)^{-1} = \left(I_n + \alpha W^{-1}\rho\right)^{-1} W^{-1} = \left(I_n - \alpha W^{-1}\rho\right)W^{-1} + O(\alpha^2),$$

where

$$W^{-1} = I_n - \frac{\gamma}{1+n\gamma} J_n,$$

and (3.2) becomes

$$v^*(t_0) = \alpha^{-1}\left\{\frac{\gamma}{1+n\gamma} + \alpha\left[1 - \frac{2\gamma}{1+n\gamma}\rho_0' 1_n + \frac{\gamma^2}{(1+n\gamma)^2} 1_n' \rho 1_n\right] + O(\alpha^2)\right\}. \tag{3.3}$$

Thus, ignoring terms of order $o(\sigma^{-2})$, G-optimality (2.2) leads to the problem of maximizing, by choice of design $\{t_1, \ldots, t_n\}$, the minimum over t_0 of

$$2(1+n\gamma)\sum_{j=1}^{n} \rho(t_0, t_j) - \gamma \sum_{i=1}^{n}\sum_{j=1}^{n} \rho(t_i, t_j). \tag{3.4}$$

This optimization problem is simplified if we regard a design as a probability measure ξ on T. Adopting the common terminology associated with this device, we speak of the *exact* problem when $n\xi(\{t\})$ is required to be integral for all $t \in T$ [as in (3.4)] and the *approximate* problem when this restriction is relaxed. Solutions to the latter are seldom directly applicable in practice; however, they are often useful for suggesting the nature of optimal exact designs and for providing bounds on the exact optimality criteria. Except for the numerical examples of Section 5, the results of this paper are for the approximate problem.

The approximate problem that arises from (3.4) is

$$\max_{\xi} \min_{t_0} \left\{2(1+\gamma^{-1}n^{-1})E_\xi \rho(t_0, X) - E_\xi \rho(X, Y)\right\}, \tag{3.5}$$

where X, Y is a sample of size 2 from ξ.

Next consider A-optimality. Return to (3.3) and take the expected value as if t_0 has distribution π. Again ignoring terms of order $o(\sigma^{-2})$, A-optimality leads to maximizing

$$2(1+n\gamma)\sum_{j=1}^{n} E_\pi \rho(\Pi, t_j) - \gamma \sum_{i=1}^{n}\sum_{j=1}^{n} \rho(t_i, t_j); \tag{3.6}$$

the associated approximate problem is

$$\max_{\xi} \left\{2(1+\gamma^{-1}n^{-1})E_{\pi,\xi}\rho(\Pi, X) - E_\xi \rho(X, Y)\right\}, \tag{3.7}$$

where X, Y is a sample of size 2 from ξ, and Π is a sample of size one from π.

For D-optimality, the generalized variance in (2.5) is

$$D = |vJ_n + \rho + \sigma^2 I_n| = \alpha^{-n}|\gamma J_n + \alpha\rho + I_n|. \tag{3.8}$$

Since $|W + \alpha\rho| = |W|\left(1 + \alpha\,\mathrm{tr}(W^{-1}\rho) + O(\alpha^2)\right)$,

$$D = \alpha^{-n}|\gamma J_n + I_n|\left\{1 + \left(n - \frac{\gamma}{1+n\gamma}1'\rho 1\right)\alpha + O(\alpha^2)\right\}. \tag{3.9}$$

Ignoring terms of order $o(\sigma^{-2})$, D-optimality requires minimizing $\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(t_i, t_j)$, and this independent of γ. In parallel with (3.5) and (3.7), the associated approximate problem is

$$\min_{\xi} E_\xi \rho(X, Y). \tag{3.10}$$

REMARK 1. γ has been fixed above, but if we take $\gamma \to 0$ and $\gamma \to \infty$, new problems can be read off from (3.4)–(3.7). In particular, as $\gamma \to 0$, (3.5) becomes

$$\max_{\xi} \min_{t_0} E_\xi \rho(t_0, X) \tag{3.11}$$

while (3.7) is

$$\max_{\xi} E_{\pi,\xi}\rho(\Pi, X). \tag{3.12}$$

Note that (3.11) is the problem of finding the optimal strategy for the maximizing player in a zero-sum game with payoff kernel ρ. On the other hand (3.12) calls for the determination of proper play given knowledge of the other player's strategy. The last problem is essentially trivial and may be solved with a one-point design measure. Here we see a connection with earlier Bayesian design results which quite commonly produce one-point measures [see Section 3.6 of Bandemer, Nather and Pilz (1987) or El-Krunz and Studden (1991) for information on this].

REMARK 2. Consider instead the model $Y = Z +$ noise, that is take $v = 0$. It is a point of some curiosity that the asymptotic problems produced by letting σ^2 go to infinity are precisely those given already in (3.10), (3.11) and (3.12), but with ρ replaced by ρ^2. Inasmuch as ρ^2 is a correlation function whenever ρ is, the basic nature of the problems is unchanged.

REMARK 3. We shall demonstrate the equivalence of (2.4) and (2.5) when T is a finite set, and without regard to any asymptotics. Let a design call for n_i observations at t_i, $i = 1,\dots,r$, and no observation at the remaining sites $t_{r+1},\dots,t_m$; $n_1 + \cdots + n_r = n$. Consider the generalized variance of observation and signal variables in the order

$$\{Y_{t_1,1},\dots,Y_{t_1,n_1},\dots,Y_{t_r,1},\dots,Y_{t_r,n_r},\beta + Z_{t_1},\dots,\beta + Z_{t_r},\beta + Z_{t_{r+1}},\dots,\beta + Z_{t_m}\},$$

where now the t_i's are distinct and $Y_{t_i,j}$ is the jth observation at t_i. The first n_1 rows of this determinant are all associated with the observations at the first design site; the jth such row is

$$\begin{gathered}(v+1)1'_{n_1} + \delta'_j\sigma^2,\dots,\big(v+\rho(t_1,t_r)\big)1'_{n_r},\\ v+1,\dots,v+\rho(t_1,t_r),\dots,v+\rho(t_1,t_m),\end{gathered} \tag{3.13}$$

where δ_j is the n_1-vector with 1 in position j and 0's everywhere else. The $(n+1)$th row, which corresponds to the signal at the same (first) design site, is identical, except that it lacks the term in σ^2. Subtract this row from (3.13), leaving the latter with σ^2 in the jth position and 0's everywhere else; the determinant is unchanged. Do this for all j's at the first site, then carry out the analogous operation at all design sites, each time subtracting the row corresponding to the signal at that site from all the rows corresponding to the observations at that site. Then the determinant has the form

$$\begin{vmatrix} I_n\sigma^2 & 0 \\ A & \rho_{\text{sig}} \end{vmatrix},$$

where ρ_{sig} is the covariance matrix for the m distinct signal variables. Thus

$$\text{GenVar(observations, signals)} = \sigma^{2n}|\rho_{\text{sig}}|, \tag{3.14}$$

which is constant over design. As the constant is the product of (2.4) and (2.5), the conclusion is at hand.

REMARK 4. The assumption that the signal has constant variance over T reflects the thought that our understanding (or ignorance) of the signal is site-independent. This allows as priors the random trigonometric polynomials which appear in the spectral representation of a stationary process with finite spectrum, for instance. At the same time, the more usual random polynomials of some fixed degree d do not fit in. On the other hand it is not difficult to produce criteria analogous to those given at (3.5), (3.7) and (3.10) that will account for this added complication.

4. Examples of optimum approximate designs. The intention in this section is to suggest some qualitative results through a sequence of remarks about, and examples of, optimum approximate designs. Completeness is not a goal, but we hope to convey some of the role played by the smoothness of Z, along with the nature and increase of the difficulties in solving problems as one proceeds from D- to A- to G-optimality.

A condition for D-optimality. Start with (3.10) in the form of minimizing

$$D(\xi) = \int_T \int_T \rho(s,t)\xi(ds)\xi(dt), \tag{4.1}$$

and note that positive definiteness of ρ makes D convex in ξ. Standard perturbation arguments then show ξ^* to be D-optimum if and only if

$$H(s) = \int_T \rho(s,t)\xi^*(dt) = h \quad \text{for all } s \in \operatorname{Supp}(\xi^*), \tag{4.2}$$

where h is the minimum of $H(s)$ on T.

A class of signals. Our main interest is infinite-dimensional processes (notwithstanding Example 1), so one requires a large enough design space T. A study of the features stemming from finite-dimensional problems will be reported on elsewhere. Thus for a standard setting we take $\{Z\}$ to be a stationary process on $[-1,1]$ with correlation function κ, that is, $\rho(t,s) = \kappa(s-t)$. Then κ has the representation

$$\kappa(u) = \int_{-\infty}^{\infty} \cos \lambda u \, F(d\lambda).$$

for a symmetric spectral distribution F. (We rule out those F's with atoms at the origin, as such signals would be partially confounded with the random constant β in our model). In this setting, if ξ on $[-1,1]$ is the distribution of a random variable V, take ξ^- to be the distribution of $-V$. It then follows that $D(\xi) = D(\xi^-)$ and, invoking convexity, that D is minimized at a ξ which is symmetric about 0.

Discrete designs. From (4.2), we see that if

$$1 + \kappa(2) \le \kappa(1 - s) + \kappa(1 + s), \tag{4.3}$$

for $|s| \le 1$, then the symmetric design having all its weight on ± 1 is D-optimum. This holds, for example, if F is supported on $[-\pi/2, \pi/2]$ or if κ is concave on $[-2, 2]$. If the spectral distribution has a second moment, κ is concave on a suitably small interval about 0, so if T is a small enough interval in R^1, a D-optimum design exists with all its weight on the boundary of T.

EXAMPLE 1. Take $\kappa(u) = \cos(\omega u), \omega > 0$. The process Z is but two-dimensional here and one might expect a design supported on two points. Rather directly from (4.3) one finds that if $\omega \le \pi/2$, the D-optimum design puts weight $\frac{1}{2}$ on ± 1. On the other hand, for $\omega > \pi/2$, the optimum design puts weight $\frac{1}{2}$ on $\pm\pi/(2\omega)$.

The case of a finite number of frequencies, $\kappa(u) = \sum_j A_j \cos(\omega_j u)$ grows more complicated, but we recall that if $|\omega_j| \le \pi/2$ for all j, the support of F ensures that (4.3) holds, and the D-optimum design again puts weight $\frac{1}{2}$ on ± 1.

EXAMPLE 2. An analytically convenient correlation corresponding to a (very) smooth, yet infinite-dimensional, Z is $\kappa(u) = \exp(-\theta u^2)$. If $\theta = \frac{1}{8}, \kappa$ is concave on $[-2, 2]$ and it follows that the D-optimum design puts all its weight on ± 1. This design remains optimum for $\theta = \frac{1}{2}$ although the correlation is now convex on $[1, 2]$; one checks (4.3). On the other hand, for $\theta = 1$, the D-optimum design is supported on 0 and ± 1, with a weight of about 0.2 at 0.

For a fixed θ, D-optimum designs here have finite support. As θ goes to infinity, this support grows dense in $[-1, 1]$. The first of these points can be argued by contradiction as follows. Suppose a D-optimum ξ^* has a cluster point in its support. Then along a convergent sequence $\{s_j\}$,

$$\begin{aligned}\exp(\theta s_j^2) H(s_j) &= \exp(\theta s_j^2) \int_{-1}^{1} \exp[-\theta(s_j - t)^2] \xi^*(dt) \\ &= \int_{-1}^{1} \exp[2\theta s_j t - \theta t^2] \xi^*(dt) = h \exp(\theta s_j^2),\end{aligned}$$

$h = \min H(s)$. Now given the equality of two Laplace transforms along $\{s_j\}$, the corresponding measures should agree. However, observe that one of them is supported on $[-1, 1]$ while the other is proportional to a normal distribution.

To understand the increasing density of D-optimum measure support points as $\theta \to \infty$, let $\theta_r \to \infty$ with ξ_r^* being D-optimum and $\text{Supp}(\xi_r^*) = \{t_{j,r}\}$. Suppress the index r and suppose one has $t_{j_0} \le a < b \le t_{j_0+1}$. (This can be arranged along a sequence of θ's approaching infinity, provided the supports are not becoming dense.) Then

$$2\min H(s) = H(t_{j_0}) + H(t_{j_0+1}) = \sum \xi_j \left[\exp[-\theta(t_j - t_{j_0})^2] + \exp[-\theta(t_j - t_{j_0+1})^2]\right]$$
$$\leq 2H\left(\frac{t_{j_0} + t_{j_0+1}}{2}\right) = 2\sum \xi_j \left[\exp\left[-\theta\left(t_j - \frac{t_{j_0} + t_{j_0+1}}{2}\right)^2\right]\right].$$

This last is contradicted by the fact that, for each individual term,

$$\exp[-\theta(t_j - t_{j_0})^2] + \exp[-\theta(t_j - t_{j_0+1})^2] > 2\exp\left[-\theta\left(t_j - \frac{t_{j_0} + t_{j_0+1}}{2}\right)^2\right],$$

for all $-1 \leq t_j \leq a$ or $b \leq t_j \leq 1$, provided θ is large enough.

Continuous designs.

EXAMPLE 3. If $\kappa(u) = \exp(-\theta|u|)$, then $H(s) = \int \exp(-\theta|s-t|)\xi^*(dt)$, which is constant on $[-1, 1]$ if

$$\xi^* = \frac{\theta}{\theta+1}U(-1,1) + \frac{1}{2(1+\theta)}(\delta_{-1} + \delta_1),$$

where U denotes the uniform measure. Specifically,

$$\begin{aligned}
H(s) &= \int_{-1}^{1} \frac{\theta}{2(1+\theta)}\exp(-\theta|s-t|)\, dt \\
&\quad + \frac{1}{2(1+\theta)}\Big(\exp(-\theta|1-s|) + \exp(-\theta|1+s|)\Big) \\
&= \frac{2 - \Big(\exp[-\theta(1-s)] + \exp[-\theta(1+s)]\Big)}{2(1+\theta)} \\
&\quad + \frac{1}{2(1+\theta)}\Big(\exp[-\theta(1-s)] + \exp[-\theta(1+s)]\Big) \\
&= \frac{1}{1+\theta}.
\end{aligned}$$

Note that, as $\theta \to \infty$, less weight is assigned to the boundary. A similar solution applies when $\kappa(u) = (1-\theta|u|)^2, 0 < \theta \leq 1$. Take

$$\xi^* = \theta U(-1,1) + \left(\frac{1}{2} - \frac{\theta}{2}\right)(\delta_{-1} + \delta_1)$$

and check that $H(s)$ is constant on $[-1, 1]$. For the (unsquared) linear correlation $\kappa(u) = 1 - \theta|u|$, (4.3) is satisfied, so the symmetric design on ± 1 is optimum.

The correlations in Example 3 are continuously differentiable off the origin with $\kappa'^{-}(0)-\kappa'^{+}(0) > 0$. When this is so, D-optimum designs do not have isolated atoms in the open interval $(-1, 1)$. For suppose a D-optimum ξ^* has isolated atoms with mass w^* at $\pm s^*, 0 \le s^* < 1$. Then

$$H(s) = \int_{\mathrm{Supp}(\xi^*)-\{\pm s^*\}} \kappa(s-t)\xi^*(dt) + w^*\kappa(s^*-s) + w^*\kappa(s^*+s)$$

has a local minimum at $s = s^*$ according to (4.2). On the other hand $H'^{-}(s^*) - H'^{+}(s^*) = w^*(\kappa'^{-}(0)-\kappa'^{+}(0)) > 0$, and this is inconsistent with such a minimum. Smoothness off the origin is necessary for this result to hold, for consider $\kappa(s) = (1-|s|)_+$. Here the D-optimum design is $\frac{1}{3}(\delta_0 + \delta_{-1} + \delta_1)$ because

$$H(s) = \tfrac{1}{3}\kappa(1-s) + \tfrac{1}{3}\kappa(s) + \tfrac{1}{3}\kappa(1+s) \equiv \tfrac{1}{3},$$

so we see that 0 is an isolated atom.

Product designs. In suitable circumstances D-optimal designs in "higher dimensions" can be obtained as products of marginal designs. Suppose $T = T_0 \times T_1$ and that the correlation between $Z_{s,s'}$ and $Z_{t,t'}$ has the product form $G(s,t)\Gamma(s',t')$, where G and Γ are correlation functions. If ξ_0^* is D-optimum on T_0 under G, and if ξ_1^* is D-optimum on T_1 under Γ, then $H_0(s) = \int G(s,t)\xi_0^*(dt) \ge h_0$ with equality on $\mathrm{Supp}(\xi_0^*)$ and $H_1(s') = \int \Gamma(s',t')\xi_1^*(dt') \ge h_1$ with equality on $\mathrm{Supp}(\xi_1^*)$. The product measure $\xi^* = \xi_0^* \times \xi_1^*$ is then D-optimum for $G \cdot \Gamma$ since

$$H((s,s')) = \int_T G(s,t)\Gamma(s',t')\xi^* d(t,t') = \int_{T_1}\int_{T_0} G(s,t)\Gamma(s',t')\xi_0^*(dt)\xi_1^*(dt') \ge h_0 \cdot h_1,$$

with equality on $\mathrm{Supp}(\xi^*)$. The D-optimality of these product designs contrasts with the poor behavior of product designs in contexts where the problem is to predict integrals of Z [see, e.g., Ylvisaker (1975)].

EXAMPLE 4. If $T = [-1,1]^2$ and $\rho((s,s'),(t,t')) = \exp[-(s-t)^2 - (s'-t')^2]$ then Example 2 with $\theta = 1$ leads to a D-optimum design concentrated on nine points—the origin, four corner points and four edge midpoints. If $\rho((s,s'),(t,t')) = \exp(-|s-t| - |s'-t'|)$, Example 3 with $\theta = 1$ gives a D-optimum design that is uniform over the square with probability $\frac{1}{4}$, is uniform along each edge with probability $\frac{1}{8}$, and has discrete weight $\frac{1}{16}$ on each of four corners.

Scaling. In Examples 2 and 3, as the "scale" parameter θ becomes larger, the correlation function becomes weaker and the D-optimum design becomes more diffuse, with less weight at the boundary. Here we explore this kind of behavior more generally. Consider a power family of correlation functions $\rho_\theta = \rho^\theta$ for positive integers θ, where ρ is a given correlation function; ρ^θ is necessarily a correlation function also. To be short and specific, we will take T to be finite, ρ to be nondegenerate and positive on $T \times T$, and argue that the uniform distribution on T is asymptotically D-optimum as $\theta \to \infty$. To see this, let ξ be any nonuniform design and let Card$(T) = m$. Then

$$\begin{aligned}\frac{\sum_s \sum_t \rho^\theta(s,t)\xi(s)\xi(t)}{\sum_s \sum_t \rho^\theta(s,t)/m^2} &\geq \frac{\sum_s \xi^2(s)}{\sum_s \sum_t \rho^\theta(s,t)/m^2} \\ &= \frac{1/m + \sum_s \left(\xi(s) - 1/m\right)^2}{\sum_s \sum_t \rho^\theta(s,t)/m^2} = \frac{1 + m\sum_s (\xi(s) - 1/m)^2}{1 + \sum \sum_{s \neq t} \rho^\theta(s,t)/m} \\ &\to 1 + m \sum_s \left(\xi(s) - \frac{1}{m}\right)^2 \quad \text{as } \theta \to \infty.\end{aligned}$$

Thus, for θ sufficiently large, ξ is worse than the uniform design.

Equivalence and G-optimality. In general, D- and G-optimum designs are different. However, if a D-optimum design ξ^* has Supp$(\xi^*) = T$, then ξ^* is G-optimum. This remark covers both (3.5) and (3.11). The result follows by considering the zero-sum game on $T \times T$ with payoff kernel ρ: If ξ^* is D-optimum and Supp$(\xi^*) = T$, then (4.2) implies that ξ^* is the optimal decision measure for both players, and the value of the game is h. In particular, $\min_{t_0} E_\xi \rho(t_0, X) \leq \min_{t_0} E_{\xi^*} \rho(t_0, X)$, where t_0 is the choice of the minimizing player. Then, if ξ is "G-better" than ξ^* under (3.5),

$$\begin{aligned}\min_{t_0} \left(\omega E_\xi \rho(t_0, X) - E_\xi \rho(X, Y)\right) &> \min_{t_0} \left(\omega E_{\xi^*} \rho(t_0, X) - E_{\xi^*} \rho(X, Y)\right) \\ &\geq \min_{t_0} \omega E_{\xi^*} \rho(t_0, X) - E_\xi \rho(X, Y),\end{aligned}$$

where $\omega = 2(1 + \gamma^{-1} n^{-1})$. This implies that $\min_{t_0} E\xi\rho(t_0, X) > \min_{t_0} E_{\xi^*} \rho(t_0, X)$, which contradicts the maximin nature of ξ^* noted above. A similar argument holds for G-optimality under (3.11).

This result can be used in Example 3 and the latter part of Example 4. Note that there is no dependence on γ. However, dependence on γ does emerge in the study of A-optimality (and for G-optimality in situations where G-and D-optimality do not agree). We will make some comments about A-optimality next. As to G-optimality, it is evident from (3.11) and, all the more, from (3.5) that one meets the usual difficulties connected with the solution of two-person zero-sum games. General problems in this area are known to be intractable, and we have nothing new to offer here.

A-optimality conditions. Here is a brief look at some features of approximate A-optimum designs. As above, let $\omega = 2(1 + \gamma^{-1} n^{-1})$. As a first case take $\gamma = \infty$ or, equivalently, $\omega = 2$. Refer to (3.7) and observe that $\xi^* = \pi$ is A-optimal because

$$
\begin{aligned}
2E_{\pi,\xi}\rho(\Pi,X) - E_{\xi}\rho(X,Y) &= 2\iint \rho(s,t)\xi(ds)\pi(dt) - \iint \rho(s,t)\xi(ds)\xi(dt) \\
&\quad - \iint \rho(s,t)\pi(ds)\pi(dt) + \iint \rho(s,t)\pi(ds)\pi(dt) \\
&= \iint \rho(s,t)\pi(ds)\pi(dt) \\
&\quad - \iint \rho(s,t)(\xi-\pi)(ds)(\xi-\pi)(dt) \\
&\le \iint \rho(s,t)\pi(ds)\pi(dt),
\end{aligned}
\tag{4.4}
$$

with equality for $\xi = \pi$. The result is intuitive enough. For a stationary correlation ρ on $[-1,1]$, and the average posterior variance criterion, the optimum design is uniform on $[-1,1]$—implemented through equal spacing, say. For a weighted average posterior variance criterion, the optimum design tracks the weighting directly. In another direction, if prediction at just a finite number of locations is important, attention is focused on these points alone. This is reminiscent of phenomena in Sacks and Ylvisaker (1985).

At the other extreme, $\omega = \infty$ or, equivalently, $\gamma = 0$, there is the problem of determining the maximizer's strategy for the game with kernel ρ and with knowledge of the minimizer's strategy π [see (3.12)]. Here a design supported entirely by one point t^* suffices, that is, choose t^* to maximize $E_{\pi}\rho(\Pi,t)$.

In the middle range, $2 < \omega < \infty$ perturbation techniques yield a condition like (4.2):

$$
H_{\omega}(s) = \int \rho(s,t)(2\xi^* - \omega\pi)(dt) = h_{\omega} \quad \text{on } \mathrm{Supp}(\xi^*),
\tag{4.5}
$$

where h_{ω} is the minimum of $H_{\omega}(s)$ on T.

EXAMPLE 5. Let $\rho(s,t) = \exp(-|s-t|)$ on $[-1,1]$ and let π be uniform on $[-1,1]$. For convenience, let $\widehat{\xi}(s) = \int \rho(s,t)\xi(dt)$ and $\widehat{\pi}(s) = \int \rho(s,t)\pi(dt)$. As remarked earlier, if $\omega = 2$, then $\xi^* = \pi$, and if $\omega = \infty$, then ξ^* is concentrated at one point, the origin, because

$$
\widehat{\pi}(s) = \tfrac{1}{2}\int_{-1}^{1} e^{-|s-t|}\,dt = 1 - \tfrac{1}{2}e^{-1}(e^{s} + e^{-s})
$$

has a maximum there. For $2 < \omega < \infty$ we attempt to satisfy (4.5). Take ξ to be uniform on $[-a,a]$ with a to be determined. Then

$$
\widehat{\xi}(s) = \begin{cases} \dfrac{1}{a} - \dfrac{e^{-a}}{2a}(e^{s} + e^{-s}), & s \in [-a,a], \\[2ex] \dfrac{e^{s}}{2a}(e^{a} - e^{-a}), & s \in [-1,-a). \end{cases}
$$

If $\hat{\xi}$ is to be $\hat{\pi} \cdot \omega/2 + c$ on $[-a, a]$ where $c = h_\omega/2$, then

$$\frac{\omega}{4} e^{-1} = \frac{e^{-a}}{2a}, \tag{4.6a}$$

$$\frac{1}{a} = \frac{\omega}{2} + c. \tag{4.6b}$$

Consider (4.6a) as an equation for a with fixed ω. The left-hand side is greater than $e^{-1}/2$ while the right-hand side decreases from ∞ to $e^{-1}/2$ as a increases from 0 to 1. Hence we can find $a = a(\omega)$ and regard (4.6a) as satisfied. (For example, if $\omega = 4, a = 0.685$, and if $\omega = 8, a = 0.4384$.) Then take c according to (4.6b). What remains then is to show that $\hat{\xi} \geq (\omega/2)\hat{\pi} + c$ on $s \in [-1, -a)$. This comes down to

$$\frac{e^s}{2a}(e^a - e^{-a}) \geq \frac{e^{1-a}}{a}\left(1 - \frac{1}{2} e^{-1}(e^s + e^{-s})\right) + \frac{1}{a}(1 - e^{1-a}) = \frac{1}{a} - \frac{e^{-a}}{2a}(e^s + e^{-s}),$$

but this holds because it is equivalent to $\exp(s + a) + \exp(-a - s) \geq 2$.

These designs are in contrast with the D- (and G) -optimal designs. Recall from Example 3 that the corresponding D-optimum design is $\frac{1}{2}U(-1, 1) + \frac{1}{4}(\delta_1 + \delta_{-1})$ and, further, that this is G-optimum as well because of the full support feature.

5. Some numerical comparisons. In this section we note some numerical results that compare the behavior of exact designs obtained using the asymptotic criteria with that of exact designs that are optimum for the original (nonasymptotic) criteria. We only report results for D-optimality. Following Section 4, we chose four cases, with dimension (k) 1 or 2 (see Table 1).We evaluate the efficiency of the asymptotic designs here by using Lindley's (1956) measure of the amount of information about a multidimensional unknown x provided by an experiment. This is the change in Shannon's entropy (i.e., prior entropy minus posterior entropy) for the random variable X that is used to represent knowledge about x. If X is an m-dimensional Gaussian variable, this is $(m/2)(\log|\Sigma_1| - \log|\Sigma_2|)$, where Σ_1 and Σ_2 are, respectively, the prior and posterior covariance matrices for X. We use this to compare the D-optimum exact design S [calculated by maximizing D in (3.8)] with the asymptotically optimum exact design S^* [calculated to minimize $\Sigma_{i=1}^n \Sigma_{j=1}^n \rho(t_i, t_j)$]. For the m

TABLE 1

Case	k	n	Correlation	Reference
1	1	4	$\rho(s,t) = \exp\left[-(s-t)^2\right]$	Example 2
2	1	4	$\rho(s,t) = \exp\left(-\lvert s-t\rvert\right)$	Example 4
3	2	11	$\rho\left((s,s'), (t,t')\right) = \exp\left[-(s-t)^2 - (s'-t')^2\right]$	Example 5
4	2	11	$\rho\left((s,s'), (t,t')\right) = \exp\left[-\lvert s-t\rvert - \lvert s'-t'\rvert\right]$	Example 5

TABLE 2

σ^2	γ	$D(S^*)/D(S)$	Eff(S^*)
0.1	0.1	0.44	0.90
	1	0.44	0.90
	10	0.44	0.91
1	0.1	0.92	0.97
	1	0.93	0.98
	10	0.94	0.99
10	0.1	1.00	1.00
	1	1.00	1.00
	10	1.00	1.00

signal variables, the information provided by a design is $(m/2)(-n\log\sigma^2+\log D)$. This can be seen by writing (3.14) as GenVar(observations) $\times$ GenVar(signal | observations) and applying the definition of information. The information provided by the asymptotically optimum design S^* relative to the optimum design S is therefore

$$\text{Eff}(S^*) = \frac{-n\log\sigma^2 + \log D(S^*)}{-n\log\sigma^2 + \log D(S)}.$$

CASE 1 [$k = 1$, $n = 4$, $\rho(s,t) = \exp(-(s-t)^2)$]. (See Table 2.) Here the asymptotic design $S^* = (-1.0, -0.36, 1.0, 1.0)$. For $\sigma^2 = 0.1$ and $\sigma^2 = 1$, S is symmetric about 0, is very nearly equispaced and includes the two endpoints; the effect of γ is negligible. For $\sigma^2 = 10$, S is similar to the asymptotic design, but with the single interior point closer to 0.

CASE 2 [$k = 1$, $n = 4$, $\rho(s,t) = \exp(-|s-t|)$]. (See Table 3.) Here the asymptotic design $S^* = (-1.0, -0.49, 0.49, 1.0)$. In all nine cases, S is similar to this, in that it is symmetric about 0 and includes the two endpoints. The distance between the two interior points increases from about 0.67 to about 0.89 as σ^2 increases from 0.1 to 10; the effect of γ is negligible.

TABLE 3

σ^2	γ	$D(S^*)/D(S)$	Eff(S^*)
0.1	0.1	0.91	0.99
	1	0.91	0.99
	10	0.91	0.99
1	0.1	0.98	0.99
	1	0.98	1.00
	10	0.98	1.00
10	0.1	1.00	1.00
	1	1.00	1.00
	10	1.00	1.00

TABLE 4

σ^2	γ	$D(S^*)/D(S)$	Eff(S^*)
0.1	0.1	0.06	0.88
	1	0.06	0.88
	10	0.06	0.88
1	0.1	0.74	0.96
	1	0.76	0.97
	10	0.76	0.97
10	0.1	1.00	1.00
	1	1.00	1.00
	10	1.00	1.00

CASE 3 [$k = 2, n = 11, \rho((s,s'),(t,t')) = \exp(-(s-t)^2 - (s'-t')^2)$]. (See Table 4.) Here S^* has two replicates at each of three corners, one point at the fourth corner and a point near the middle of each edge. It is symmetric about the diagonal line passing through the unreplicated corner. (See Figure 1.) For the nonasymptotic *D-optimal* designs, one sees no replication except when $\sigma^2 = 10$. For $\sigma^2 = 0.1$, the *D-optimal* design S is essentially the same for each γ. There are four corner points, a point near the middle of each edge and a nearly equilateral triangle in the interior, these designs are symmetric about one of the diagonals. For $\sigma^2 = 1$, γ again has little effect. In each case, S

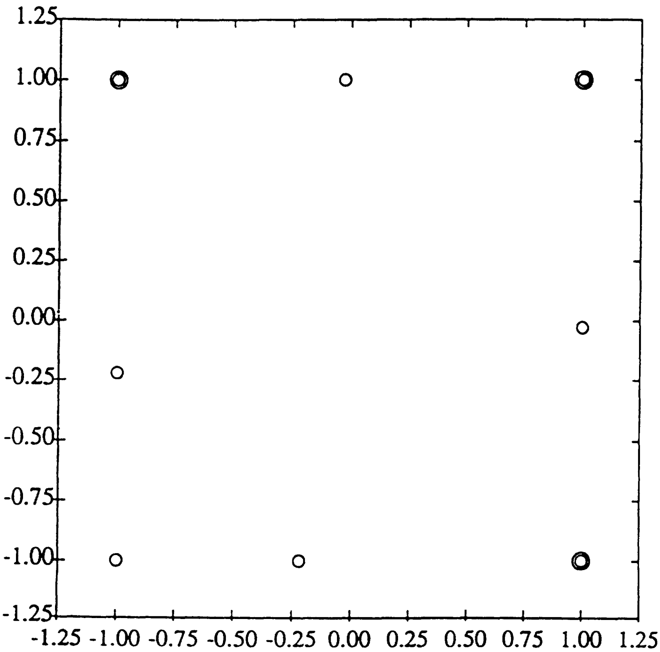

FIG. 1. *An asymptotic D-optimal* 11*-run design on* $[-1,1]^2$, *where* $\rho((s,s'),(t,t')) = \exp[-(s-t)^2 - (s'-t')^2]$. *Points at which there are replicates are indicated by the double circles.*

TABLE 5

σ^2	γ	$D(S^*)/D(S)$	Eff(S^*)
0.1	0.1	0.79	0.99
	1	0.79	0.99
	10	0.79	0.99
1	0.1	0.95	0.99
	1	0.95	0.99
	10	0.95	1.00
10	0.1	1.00	1.00
	1	1.00	1.00
	10	1.00	1.00

is symmetric about both coordinate axes. There are four corner points, a center point, a point at the center of two opposing edges and two points on each of the remaining two edges, each at a distance of about one-third from the middle of the edge. For $\sigma^2 = 10$ and $\gamma = 0.1$ or 1, S is symmetric about both diagonals. There is a center point, two replicated corners, two unreplicated corners and a point near the middle of each edge. In the last case ($\sigma^2 = 10$, $\gamma = 10$), S is practically identical to the asymptotic design S^*.

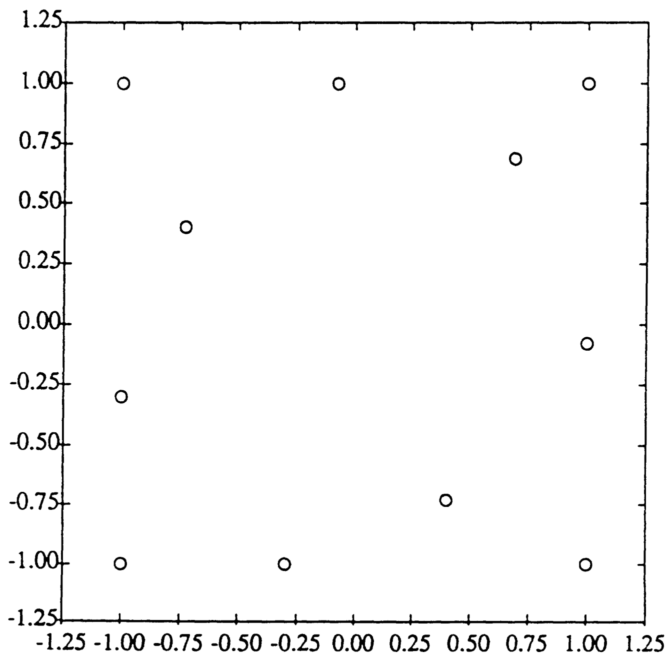

FIG. 2. *An asymptotic D-optimal* 11-*run design on* $[-1, 1]^2$, *where* $\rho((s, s'), (t, t')) = \exp(-|s - t| - |s' - t'|)$.

CASE 4 [$k = 2, n = 11, \rho((s,s'),(t,t')) = \exp(-|s-t| - |s'-t'|)$]. (See Table 5.) As one might expect from the examples of Section 4, the designs here favor the interior more than those for the Gaussian correlation. There are no replicates in any of these designs. The asymptotic design S^* has four corner points, a point near the center of each edge and a large, nearly equilateral, triangle in the interior (Figure 2). It is symmetric about one of the diagonals, as are all of the nonasymptotic D-optimal designs. For $\sigma^2 = 0.1$, S includes five interior points, four corner points and the edge midpoints of two adjoining edges; γ has no discernible effect. Very similar designs were obtained for $\sigma^2 = 1$. For $\sigma^2 = 10$, S looks like the asymptotic design in each instance, expect that the interior triangle is slightly shrunken.

Finding optimum designs is a major computational problem, particularly because of the existence of multiple locally optimum designs, and this has limited our attention here to these few simple cases. Even so, we cannot guarantee that we have found the best design in every case. It is nevertheless encouraging that the asymptotic designs constructed here, which do not depend on γ or σ^2, appear to be adequate over fairly wide ranges of these parameters.

Acknowledgments. We thank Bidan Huang and Robert Buck for making the computations reported on in Section 5. We also thank a referee for several suggestions which have led to improvements in the paper.

REFERENCES

BANDEMER, H., NATHER, W. and PILZ, J. (1987). Once more: optimal experimental design for regression models (with discussion). *Statistics* **18** 171–217.

BICKEL, P. J. and HERZBERG, A. M. (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Ann. Statist.* **7** 77–95.

CHALONER, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* **12** 283–300.

EL-KRUNZ, S. M. and STUDDEN, W. J. (1991). Bayesian optimal design for linear regression models. *Ann. Statist.* **19** 2183–2208.

JOHNSON, M., MOORE, L. and YLVISAKER, D. (1990). Minimax and maximin distance designs. *J. Statist. Plann. Inference* **26** 131–148.

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **12** 363–366.

LIM, Y., SACKS, J., STUDDEN, W. and WELCH, W. (1988). Optimum designs for highly correlated computer experiments. Unpublished manuscript.

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

O'HAGAN, A. (1978). Curve fitting and optimal design for prediction (with discussion). *J. Roy. Statist. Soc. Ser. B* **40** 1–42.

PILZ, J. (1983). *Bayesian Estimation and Experimental Design in Linear Regression Models. Teubner-Texte zur Math.* **55**. Teubner, Leipzig.

SACKS, J., WELCH, W., MITCHELL, T. and WYNN, H. (1989). Design and analysis of computer experiments (with discussion). *Statist. Sci.* **4** 409–423.

SACKS, J. and YLVISAKER, D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.* **37** 66–89.

SACKS, J. and YLVISAKER, D. (1985). Model robust design in regression: Bayes theory. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. LeCam and R. Olshen eds.) **2** 667–679. Wadsworth, Monterey, CA.

STEINBERG, D. (1985). Model robust response surface designs: scaling two-level factorials. *Biometrica* **72** 219–230.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

YLVISAKER, D. (1975). Designs on random fields. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.). North-Holland, Amsterdam.

TOBY MITCHELL
ENGINEERING PHYSICS
AND MATHEMATICS DIVISION
OAK RIDGE NATIONAL LABORATORY
OAK RIDGE, TENNESSEE 37831

JEROME SACKS
NATIONAL INSTITUTE
OF STATISTICAL SCIENCES
RESEARCH TRIANGLE PARK,
NORTH CAROLINA 27709

DONALD YLVISAKER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024