Statistica Sinica **5**(1995), 559-573

TWO-LEVEL FRACTIONAL FACTORIALS AND BAYESIAN PREDICTION

Toby J. Mitchell, Max D. Morris and Donald Ylvisaker*

*Oak Ridge National Laboratory and University of California**

Abstract. The paper considers the problem of design for prediction of a deterministic response function x over a domain T. A Bayesian approach is used, where the random function that represents prior uncertainty about x is a stationary Gaussian stochastic process X. Here $T = \{-1, 1\}^k$, the designs considered are fractional factorials, and the objective is to optimize the choice of design with respect to some criterion. The structure of stationary and of isotropic processes on T is discussed, along with the conditioning of such a process based on observation at a fractional factorial design. There are useful regularities in this, together with workable criteria on the prediction of interactions and on the prediction of unobserved values of the process.

Key words and phrases: Bayesian prediction, computer experiments, fractional factorial designs, stationary processes, two-level factors.

1. Introduction

We consider computer experiments in which a numerical response is associated with the setting of k inputs or factors. As would often be the situation with physical experiments, we are concerned with cases in which the number of observations may not be large relative to k, which may be very large. Our view of inference here falls into the general framework of the review paper of Sacks, Welch, Mitchell and Wynn (1989).

The analysis will be Bayesian in that uncertainty about the response is represented by a random process on T, the set of interesting k-vector inputs. The analysis is straightforward: for a given prior one performs an updating after sparse observation in T, and this results in predictions of, for example, responses at unobserved sites and uncertainty assessments about these predictions. This approach is outlined in Currin, Mitchell, Morris and Ylvisaker (1991) and, without computer experiments in the foreground, one can look back to the work of Kimeldorf and Wahba (1970).

In fact our primary interest lies in the Bayesian design problem. Given scarce resources, where should one observe to obtain the highest efficiency in prediction? The general area of Bayesian design for prediction is surveyed in Ylvisaker (1987)

when observation is made with or without error. (We make some connection in Section 3 between these two contexts for our set-up.) O'Hagan (1978) is an early Bayesian design paper that invokes "nonparametric" priors in the presence of observational error.

Specific design issues have been addressed in recent literature, but real answers are thus far incomplete. When T is scaled to the unit cube, Latin hypercube samples, introduced by McKay, Conover and Beckman (1979), have the intuitively appealing property of uniform 1-dimensional projections. Tang (1993) has suggested designs which combine Latin hypercube and orthogonal array structures, resulting in interesting projections in small-dimensional subspaces. Morris and Mitchell (1995) have constructed designs which, with respect to a certain criterion, are optimal for prediction within the class of designs which are constructed as Latin hypercubes.

On the other hand, it is not clear that having a large number of values represented for each factor is necessary or even useful in computer experimentation. Sacks et al. (1989) used a predictive integrated mean square error criterion to generate a 16 run design, for 6 factors, which very closely resembles an irregular 2-level fraction with center point. Easterling (1989), in his review of that article, argued for geometrically simpler designs such as fractional factorials. Johnson, Moore and Ylvisaker (1990) developed connections between geometric and statistical properties of designs. In particular, the class of "maximin distance" designs, which includes the two-level Plackett and Burman (1943) designs, was shown to possess a Bayesian D-optimality property for locally weak correlations.

Here we shall restrict not only designs but also inferences to $T = \{-1, 1\}^k$. Particularly when the number of observations is not large relative to k, it may often be too much to ask that a data set support good prediction across a continuum of high dimension. Early experimental considerations are more often centered in determining which factors are most important and which combinations of them may be associated with important interactions. Even in this restricted framework, there are no universally optimum designs - matters do depend on the prior selected and on the criteria of efficiency adopted.

In Section 2 we look at functions over T and at the stationary and isotropic processes that are supported there. The structure of such processes is neatly tied to prior independence of interactions and to relationships between interaction variances. Though the point is not explicitly pursued here, the material in this section is useful in setting prior processes when uncertainty about factor interactions is the information most naturally elicited. Fractional factorial designs are brought forward in Section 3 and it turns out that observing a stationary process at such a set of runs allows a simple updating mechanism. The final section

then gets closer to the actual computational problems associated with choosing a design; there, amongst other things, we examine effects of the prior and the interaction set of interest on design selection via D-optimality for a specific example of five factors.

The paper is rather self-contained in its theoretical outline, and much of what appears is not really unexpected. In particular, we carry out in some detail, and on a very specific space, a program that is promised success in the work of Yaglom (1961) dealing with stationary processes on groups. In defense of this posture, a good deal of notation is needed and there is not much extra expense in putting in some "well-known" facts. After these details are collected, computational issues would allow us to go on at considerable length. Nonetheless Section 4 is more for the sake of illustrating problems that can be handled, as concrete matters are to be addressed more fully elsewhere.

2. Preliminaries and Priors

The set of experimental "runs" is $T = \{-1, 1\}^k$ and the generic element of T is $\boldsymbol{t}$, say, with coordinates t_i, $i = 1, \ldots, k$. There are two mathematical ways to view T that will be useful. First of all, T forms a group under the direct multiplication operation: $\boldsymbol{st}$ denotes the k-vector with coordinates $s_i t_i$. In particular the unit element of the group is $\mathbf{1}$ and every element is self-inverse. Secondly T is a metric space under Hamming distance: $d(\boldsymbol{s}, \boldsymbol{t})$ is the number of coordinates in which $\boldsymbol{s}$ and $\boldsymbol{t}$ differ; variously, one can write $d(\boldsymbol{s}, \boldsymbol{t}) = (1/2) \sum |s_i - t_i|$.

We make use of the collection of all mappings from T to $\{-1, 1\}$. Our notation here goes as follows: if A is a subset of $\{1, \ldots, k\}$, let $\boldsymbol{A}$ be defined on T by $\boldsymbol{A}(\boldsymbol{s}) = \prod_{j \in A} s_j$.

Now let x be a real-valued function on T and consider its ANOVA expansion. The *interaction term associated with a subset A of* $\{1, \ldots, k\}$ is $\Gamma_A(x) = 2^{-k} \sum_{\boldsymbol{s}} \boldsymbol{A}(\boldsymbol{s}) x(\boldsymbol{s})$, and the *order* of Γ_A is the cardinality of A. It follows that x can be written

$$x(\boldsymbol{t}) = \sum_{\boldsymbol{s}} x(\boldsymbol{s}) \{2^{-k} \prod_j (1 + s_j t_j)\} = 2^{-k} \sum_{\boldsymbol{s}} x(\boldsymbol{s}) \sum_A \boldsymbol{A}(\boldsymbol{s}) \boldsymbol{A}(\boldsymbol{t}) = \sum_A \boldsymbol{A}(\boldsymbol{t}) \Gamma_A(x). \tag{2.1}$$

As far as priors over the space of functions on T are concerned, (2.1) makes clear that one can provide the joint distribution of random variables $X(\boldsymbol{t})$ for $\boldsymbol{t}$ in T and the joint distribution of the Γ's will follow, or vice versa. This will be made more explicit below for the priors of interest. All of our joint distributions for the X's (equivalently Γ's) will be taken to be normal and, in consequence, posterior predictions will be linear and residual covariance will be data-independent.

Stationary priors

Suppose X is a Gaussian process on T with mean zero. Call X *stationary* provided for any $\boldsymbol{m}$ in $\{-1,1\}^k$, $\{Z(\boldsymbol{t})\}$ is distributed as $\{X(\boldsymbol{t})\}$ where $Z(\boldsymbol{t}) = X(\boldsymbol{mt})$.

Proposition 2.1. *X is stationary if and only if the interactions $\{\Gamma_A\}$ are independent.*

Proof. Let X be stationary. Consider distinct sets A and B and, in particular, assume j is in B and not in A. Let $\boldsymbol{m}$ have coordinates 1 with the exception of the jth which is -1. If $Z(\boldsymbol{t}) = X(\boldsymbol{mt})$, for all $\boldsymbol{t}$ in T, one finds that $2^{2k}E\Gamma_A(X)\Gamma_B(X)$ is given by

$$\begin{aligned}
&E\Big[\sum_{\boldsymbol{A}(\boldsymbol{s})=1} X(\boldsymbol{s}) - \sum_{\boldsymbol{A}(\boldsymbol{s})=-1} X(\boldsymbol{s})\Big]\Big[\sum_{\boldsymbol{B}(\boldsymbol{t})=1} X(\boldsymbol{t}) - \sum_{\boldsymbol{B}(\boldsymbol{t})=-1} X(\boldsymbol{t})\Big] \\
=&E\Big[\sum_{\boldsymbol{A}(\boldsymbol{s})=1} Z(\boldsymbol{s}) - \sum_{\boldsymbol{A}(\boldsymbol{s})=-1} Z(\boldsymbol{s})\Big]\Big[\sum_{\boldsymbol{B}(\boldsymbol{t})=1} Z(\boldsymbol{t}) - \sum_{\boldsymbol{B}(\boldsymbol{t})=-1} Z(\boldsymbol{t})\Big] \\
=&E\Big[\sum_{\boldsymbol{A}(\boldsymbol{s})=1} X(\boldsymbol{s}) - \sum_{\boldsymbol{A}(\boldsymbol{s})=-1} X(\boldsymbol{s})\Big]\Big[\sum_{\boldsymbol{B}(\boldsymbol{t})=-1} X(\boldsymbol{t}) - \sum_{\boldsymbol{B}(\boldsymbol{t})=1} X(\boldsymbol{t})\Big]
\end{aligned}$$

which is the negative of $2^{2k}E\Gamma_A(X)\Gamma_B(X)$.

Conversely, let the Γ's be independent mean zero normals with variances designated by V, indexed appropriately. From (2.1) we compute

$$EX(\boldsymbol{s})X(\boldsymbol{t}) = \sum_A V_A \boldsymbol{A}(\boldsymbol{s})\boldsymbol{A}(\boldsymbol{t}) = \sum_A V_A \boldsymbol{A}(\boldsymbol{ms})\boldsymbol{A}(\boldsymbol{mt}) = EX(\boldsymbol{ms})X(\boldsymbol{mt})$$

since $\boldsymbol{m}^2 = \boldsymbol{1}$.

Here is a particular collection of stationary processes that is computationally pleasant. Return to the representation at (2.1) and endow independent Γ's with variances V that depend on the particular index set as follows. Start with a sequence of positive correlations $\rho_1, \rho_2, \ldots, \rho_k$. Take $V_A = 2^{-k}\prod_{i\notin A}(1+\rho_i)\prod_{i\in A}(1-\rho_i)$ and find

$$\begin{aligned}
EX(\boldsymbol{s})X(\boldsymbol{t}) &= \sum_A \Big[2^{-k}\prod_{i\notin A}(1+\rho_i)\prod_{i\,\in\, A}(1-\rho_i)\boldsymbol{A}(\boldsymbol{s})\boldsymbol{A}(\boldsymbol{t})\Big] \\
&= 2^{-k}\prod[(1+\rho_i)+(1-\rho_i)s_it_i] = \prod_{s\neq t}\rho_i = \prod_{\rho_i}^{\frac{1}{2}|s_i-t_i|}. \qquad (2.2)
\end{aligned}$$

Note the nested property of the variances in this construction: since the ρ_i are all positive, $A \subset B$ implies $V_B \le V_A$. That is, higher order interaction terms are assigned less variability by such priors.

The correlation functions in question have a weak sense decreasing character as well. Say that $\boldsymbol{t}$ is *between* $\boldsymbol{s}$ and $\boldsymbol{u}$ provided $d(\boldsymbol{s},\boldsymbol{u}) = d(\boldsymbol{s},\boldsymbol{t}) + d(\boldsymbol{t},\boldsymbol{u})$. Now with positive ρ's it is easy to see that for $\boldsymbol{t}$ between $\boldsymbol{s}$ and $\boldsymbol{u}$, $\mathrm{Corr}\,(X(\boldsymbol{s}), X(\boldsymbol{u})) \leqq \mathrm{Corr}\,(X(\boldsymbol{s}), X(\boldsymbol{t}))$.

Isotropic priors

Suppose again that X is a Gaussian process on T with mean value zero and covariance function R. Call X *isotropic* provided $R(\boldsymbol{s},\boldsymbol{t}) = r(d(\boldsymbol{s},\boldsymbol{t}))$ for some function r defined on $\{0, 1, \ldots, k\}$.

Proposition 2.2. *The following are equivalent.*

(a) *X is isotropic on T with some covariance function r.*

(b) (i) *X is stationary.*

(ii) *$\{X(\boldsymbol{t}), \boldsymbol{t}$ in $T\}$ is distributed as $\{X(\sigma(\boldsymbol{t})), \boldsymbol{t}$ in $T\}$ for any permutation σ of $1, \ldots, k$.*

(c) *The Γ's are independent and $V_A = V_{|A|}$.*

Proof. The implication from (a) to (b) is straightforward. To go from (b) to (c), use the independence of the Γ's from Proposition 2.1. According to (b) (ii), it follows further that the variance of an interaction term depends only on the order of that interaction.

Finally, to go from (c) to (a), represent an X as at (2.1) where the Γ's are independent normals with rth order variance V_r, say. X is then a Gaussian process and

$$E\,X(\boldsymbol{s})X(\boldsymbol{t}) = V_0 + V_1 \sum s_i t_i + V_2 \sum_{i<j} s_i t_i s_j t_j + \cdots + V_k s_1 t_1 s_2 t_2 \cdots s_k t_k. \quad (2.3)$$

That (2.3) depends only on $d(\boldsymbol{s},\boldsymbol{t})$ follows by first noting that

$$\sum s_i t_i = n - \sum |s_i - t_i| = n - 2d(\boldsymbol{s},\boldsymbol{t}).$$

Successive multiplication on both sides of this last expression by $\sum s_i t_i$, followed by a rearrangement of terms, will then show that each item on the right side of (2.3) is again a function of $d(\boldsymbol{s},\boldsymbol{t})$.

It is not so hard to characterize those functions r which can serve as covariance functions of isotropic processes on $\{-1, 1\}^k$. Indeed, let X be isotropic and have covariance function r. Take $A = \{1, \ldots, h\}$, $h = 0, 1, \ldots, k$, and compute that

$$0 \leq 2^k V_h = 2^k \mathrm{Var}(\Gamma_A) = 2^{-k} \sum_{\boldsymbol{s}} \sum_{\boldsymbol{t}} EX(\boldsymbol{s})X(\boldsymbol{t})\boldsymbol{A}(\boldsymbol{s})\boldsymbol{A}(\boldsymbol{t})$$

$$= 2^{-k} \sum_{\boldsymbol{s}} \sum_{\boldsymbol{t}} EX(\boldsymbol{1})X(\boldsymbol{t})\boldsymbol{A}(\boldsymbol{1})\boldsymbol{A}(\boldsymbol{t}) = \sum_{\boldsymbol{t}} r(d(\boldsymbol{1},\boldsymbol{t}))\boldsymbol{A}(\boldsymbol{t})$$

$$= \sum_{q \geq 0} r(q) \sum_{m \geq 0} h!/[m!(h-m)!](-1)^m (k-h)!/[(q-m)!(k-h-q+m)!]. \quad (2.4)$$

Clearly the conditions on r given at (2.4) are necessary for it to qualify as an isotropic correlation function. (They can be rephrased in different forms and one of the more interesting of these makes connection with Kravchuk polynomials (MacWilliams and Sloane (1977)). We will not trouble to spell this out further here.) Going further, these conditions turn out to be sufficient as well, (see Letac (1981) for example).

Proposition 2.3. *A necessary and sufficient condition that there exist an isotropic process X on $\{-1, 1\}^k$ with covariance function r is that* (2.4) *holds for $h = 0, 1, \ldots, k$.*

There are isotropic covariance functions r that are not decreasing on $\{1, \ldots, k\}$ and so the corresponding correlations do not decrease in the weak sense described earlier under stationary processes. This holds even if higher order interactions are assigned less variability. As an example, take $k = 3$ with $V_0 = 1/2$, $V_1 = 1/12$, $V_2 = 1/15$ and $V_3 = 1/20$. It follows by calculation that $r(0) = 1$, $r(1) = 7/15$, $r(2) = 2/5$ and $r(3) = 13/30$. A decreasing r would obtain through $k = 3$ if the V's were not only decreasing but convex as well. But then with $k = 4$,

$$r(3) - r(4) = 2V_1 - 6V_2 + 6V_3 - 2V_4 = 2(V_1 - 2V_2 + V_3) - 2(V_2 - 2V_3 + V_4)$$

and convexity is not sufficient to make this difference nonnegative.

It seems desirable in an isotropic prior both that correlations be decreasing in distance and that interaction variances be decreasing as order increases. Here, then, are conditions under which these properties obtain.

Define difference operators D^m as follows. For a sequence a_j, $j = 1, 2, \ldots, k$, let

$$D^m a_h = \sum_{j \geq 0} a_{h+j} (-1)^j \, m!/j!(m-j)! \qquad \text{for } h + m \leq k.$$

We call the sequence a_j, $j = 1, 2, \ldots, k$, *completely monotone* to order q provided

$$D^m a_h \geq 0 \qquad \text{for } m = 0, 1, \ldots, \min(q, k-h).$$

With this notation, we have the following proposition.

Proposition 2.4. *If the interaction variances V_j, $j = 0, \ldots, k$, are completely monotone to order k, then the associated isotropic process has a covariance function that is decreasing in distance.*

The proof of the proposition is rather straightforward and so is not included.

The generic isotropic correlation function is $r(d) = \rho^d$ for some positive correlation ρ (that the corresponding interaction variances are completely monotone of all orders can be checked starting from (2.2)). The relevant class is broader for finite k but a family of correlations which qualifies for all k is generated by the particular one through integration, as in $r(d) = \int \rho^d \, dF(\rho)$ for some distribution function F on $[0, 1]$. Henceforth, we implicitly consider only those isotropic priors with covariance function that decrease with distance.

3. Fractional Factorials and Conditioning

We first bring in fractional factorials. John (1971), for example, treats these matters more thoroughly; here we introduce only ideas and notation necessary to support our results. We subsequently look at the problem of conditioning on the observation of a stationary process on a fractional factorial design.

First allow that runs $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_q$ are *independent* if $\prod_j \boldsymbol{m}_j^{e_j} = \mathbf{1}$, e_j is 0 or 1 for each j, implies $e_j = 0$ for all j. Let $W_1, \ldots, W_p$ be p subsets of $\{1, 2, \ldots, k\}$, called *words*, and let ϕ_j be the indicator function of the jth word. Words $W_1, \ldots, W_p$ are said to be *independent* if the runs $2\phi_j - \mathbf{1}$ are independent, $j = 1, \ldots, p$.

We fix p independent words and specify the fractional factorial design by the "defining relation" $I = W_1 = \cdots = W_p$. This is to mean that the set of runs of the fractional factorial is

$$F = \{\boldsymbol{m} | \boldsymbol{W}_v(\boldsymbol{m}) = 1 \quad \text{for } v = 1, \ldots, p\}. \tag{3.1}$$

Observe here that $\mathbf{1}$ is necessarily in the fraction and that F is closed with respect to pointwise multiplication. In other words, F is a subgroup of T under the group operation of direct multiplication. The next result provides us with the essential information about fractional factorials – it specifies a natural order for the runs that will facilitate dealing with F.

Proposition 3.1. *If $W_1, \ldots, W_p$ are independent words, then there are independent runs $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{k-p}$ so that*

$$F = \{\boldsymbol{m} | \boldsymbol{m} = \prod_j \boldsymbol{m}_j^{e_j}, \; e_j(1 - e_j) = 0 \quad \text{all } j\}. \tag{3.2}$$

A proof is not hard, and an entertaining one is available from the authors.

Throughout the rest of the section X is to be stationary on T, it has mean zero and a covariance R. Observations on X are taken at the runs specified by a fractional factorial F.

Begin by ordering the 2^{k-p} runs of F according to the representation at (3.2):

$$\boldsymbol{m} = \boldsymbol{m}(q) = \prod_j \boldsymbol{m}_j^{e_j}$$

is the qth run if q, written in binary notation in the usual way, is $e_{k-p}e_{k-p-1}\cdots e_1$, $q = 0, 1, \ldots, 2^{k-p} - 1$. Thus the ordering begins 1, $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, $\boldsymbol{m}_1\boldsymbol{m}_2$, $\boldsymbol{m}_3$, and so on. As things proceed, we blur over any distinction between $\boldsymbol{e}_q$ as the binary expansion of the integer q and the corresponding $k-p$ vector with coordinates $e_1, \ldots, e_{k-p}$.

Our interest lies in the $2^{k-p} \times 2^{k-p}$ covariance matrix $\boldsymbol{R}$ of X at the runs in the order now specified. It has as its $(i+1, j+1)$st entry $EX(\boldsymbol{m}(i))X(\boldsymbol{m}(j)) = \boldsymbol{R}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \boldsymbol{R}(\boldsymbol{1}, \boldsymbol{e}_i \oplus \boldsymbol{e}_j)$, where $\oplus$ denotes binary addition and we are exercising the stationarity assumption. For extra clarity,

$$\boldsymbol{R}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \boldsymbol{R}\Big(\prod_h \boldsymbol{m}_h^{e_{i,h}}, \prod_h \boldsymbol{m}_h^{e_{j,h}}\Big) = \boldsymbol{R}\Big(\boldsymbol{1}, \prod_h \boldsymbol{m}_h^{e_{i,h}+e_{j,h}}\Big) = \boldsymbol{R}(\boldsymbol{1}, \boldsymbol{e}_i \oplus \boldsymbol{e}_j). \quad (3.3)$$

In fact $\boldsymbol{R}$ has a simple spectral decomposition: we exhibit a symmetric matrix $\boldsymbol{E}$ with orthogonal columns and a diagonal matrix Λ so that $\boldsymbol{R} = \boldsymbol{E}\Lambda\boldsymbol{E}$. Specifically, let $\boldsymbol{E}$ have the $(i+1, j+1)$st entry

$$\boldsymbol{E}_{i,j} = (-1)^{(\boldsymbol{e}_i, \boldsymbol{e}_j)}, \quad (3.4)$$

where $(\boldsymbol{e}_i, \boldsymbol{e}_j)$ is the ordinary inner product. Observe that $\boldsymbol{E}^2$ has $(i+1, j+1)$st entry

$$\sum_q (-1)^{(\boldsymbol{e}_i, \boldsymbol{e}_q)}(-1)^{(\boldsymbol{e}_j, \boldsymbol{e}_q)} = \sum_q (-1)^{(\boldsymbol{e}_i \oplus \boldsymbol{e}_j, \boldsymbol{e}_q)}$$

and that this sum vanishes unless $\boldsymbol{e}_i = \boldsymbol{e}_j$, when it is 2^{k-p}. Thus $\boldsymbol{E}$ is symmetric and has orthogonal columns. Next let Λ be the diagonal matrix with $\Lambda_{q+1,q+1}$ given by

$$\lambda_q = \sum_v \boldsymbol{R}(\boldsymbol{1}, \boldsymbol{e}_v)(-1)^{(\boldsymbol{e}_v, \boldsymbol{e}_q)}, \quad q = 0, 1 \ldots, 2^{k-p} - 1. \quad (3.5)$$

After these remarks the result we want follows quickly.

Proposition 3.2. *If $\boldsymbol{R}$, $\boldsymbol{E}$ and Λ are given by* (3.3)-(3.5), *then $\boldsymbol{R} = \boldsymbol{E}\Lambda\boldsymbol{E}$.*

Proof. Calculate the proposed $(i+1, j+1)$st entry of $\boldsymbol{R}$ to be

$$\sum_q \lambda_q (-1)^{(\boldsymbol{e}_i, \boldsymbol{e}_q)}(-1)^{(\boldsymbol{e}_q, \boldsymbol{e}_j)} = \sum_q \Big[\sum_v \boldsymbol{R}(\boldsymbol{1}, \boldsymbol{e}_v)(-1)^{(\boldsymbol{e}_v, \boldsymbol{e}_q)}\Big](-1)^{(\boldsymbol{e}_i \oplus \boldsymbol{e}_j, \boldsymbol{e}_q)}$$

$$= \sum_v \boldsymbol{R}(\boldsymbol{1}, \boldsymbol{e}_v) \sum_q (-1)^{(\boldsymbol{e}_i \oplus \boldsymbol{e}_j \oplus \boldsymbol{e}_v, \boldsymbol{e}_q)} = \sum_v \boldsymbol{R}(\boldsymbol{1}, \boldsymbol{e}_v)\delta_{i,j,v}, \quad (3.6)$$

where $\delta_{i,j,v} = 1$ if $\boldsymbol{e}_i \oplus \boldsymbol{e}_j \oplus \boldsymbol{e}_v = 0$, and is 0 otherwise. But then the right side of (3.6) is $\boldsymbol{R}(\mathbf{1}, \boldsymbol{e}_i \oplus \boldsymbol{e}_j) = \boldsymbol{R}(\boldsymbol{e}_i, \boldsymbol{e}_j)$, the required result.

Remark 3.1. There is no harm done to the previous proposition if we take $p = 0$. Thus the covariance matrices of stationary X over $\{-1, 1\}^k$ are simultaneously diagonalized. We have come at this from an unusual direction in order to have matters set for the process restricted to F. Variously, without deliberate construction, one resorts to the general results of Yaglom (1961) about stationary processes on (sub)groups.

Equipped with the decomposition (note that the columns of $\boldsymbol{E}$ are not of unit length so that Λ is merely proportional to the eigenvalues of the covariance matrix) one can look, for example, at the generalized variance of the process at sites left unobserved through a specific fractional factorial design. That is, one can investigate such questions as the D- or G-optimality of fractional factorials under various stationary priors. Set aside these direct problems for now and consider the benefit of the special structure evident in Proposition 3.2 when one concentrates on the prediction of *interactions*, as opposed to the prediction of *unobserved values*.

To see that one meets special features when contemplating interaction terms, return to (2.1) – one has $X(\boldsymbol{t}) = \sum_A \Gamma_A(X)\boldsymbol{A}(\boldsymbol{s})$. If X is stationary, the covariance of $\Gamma_B(X)$ with the values observed at the fractional factorial F, as indexed at (3.2), might be written as $V_B \cdot \boldsymbol{B}^*(\boldsymbol{e}_j)$, $j = 0, 1 \ldots, 2^{k-p} - 1$, where

$$\boldsymbol{B}^*(\boldsymbol{e}_j) = \boldsymbol{B}\Big(\prod_v \boldsymbol{m}_v^{e_{j,v}}\Big) = \prod_v [\boldsymbol{B}(\boldsymbol{m}_v)]^{e_{j,v}}.$$

Meanwhile the $(j+1)$st coordinate of the $(i+1)$st row of $\boldsymbol{E}$ is

$$\boldsymbol{E}_{i,j} = (-1)^{(\boldsymbol{e}_i, \boldsymbol{e}_j)} = (-1)^{\Sigma_v e_{i,v} e_{j,v}} = \prod_v [(-1)^{e_{i,v}}]^{e_{j,v}}.$$

Thus, the row vector of covariances of $\Gamma_B(X)$ with the observations at the 2^{k-p} runs of F is the ith row of $\boldsymbol{E}$, up to the factor V_B, provided

$$\boldsymbol{B}(\boldsymbol{m}_v) = (-1)^{e_{i,v}} \qquad \text{for } v = 1, 2, \ldots, k-p. \tag{3.7}$$

Moreover, (3.7) holds for some i since there are 2^{k-p} $\boldsymbol{e}_i$'s available.

We will say *the subset B of $\{1, \ldots, k\}$ maps to i*, $0 \le i \le 2^{k-p} - 1$, provided (3.7) obtains. In particular one sees that the set of 2^p words generating F map to $i = 0$. Going further, an argument dual to that required for Proposition 3.1 will show that there are 2^p subsets which map to i for each i. Then, using classical language, we say interactions $\Gamma_A(X)$ and $\Gamma_B(X)$ are *aliased* provided A and B

map to the same i. The effect of aliasing on the posterior distribution comes out in the next result.

Proposition 3.3. *If A maps to i, the posterior variance of $\Gamma_A(X)$ given observation of X on F is $V_A - V_A^2\lambda_i^{-1}$, $i = 0, 1, \ldots, 2^{k-p} - 1$. If A maps to i and B maps to j, the posterior covariance of $\Gamma_A(X)$ and $\Gamma_B(X)$ given observation of X on F is $-\delta_{i,j}V_AV_B\lambda_i^{-1}$.*

Proof. The result follows directly from Proposition 2.1, Proposition 3.2 and the nature of the mapping at (3.7).

Remark 3.2. Suppose instead that one can observe a stationary process on a fractional factorial design, but only up to an error of constant variance. The addition of a multiple of the identity to the covariance R of the process will leave the eigenvectors unchanged so that Propositions 3.2 and 3.3 still apply following minor changes.

4. Fractional Factorials and Bayesian Prediction

The present section is a brief exploration of the computational problems which are met in the choice of fractional factorial beginning from a stationary prior, and with an eye on the posterior variability of interaction terms or process values at unobserved sites. When one focuses attention on interaction terms, the design criteria utilize Proposition 3.3 quite directly. They differ from criteria now standardly associated with the posterior prediction of unobserved values. On the other hand, they have the advantage of connection with classical ways of thinking about the unknown x and there can be fewer computational difficulties.

We limit the present discussion to illustrating what we have in mind rather than aiming for any kind of completeness. A more systematic study will appear elsewhere.

Start with an X having the simplest isotropic correlation $\rho(d) = \rho^d$, ρ positive, and profess an interest in low order interactions. From (2.2), the mean (0th order) effect has variance $V = V_0 = 2^{-k}(1+\rho)^k$, the main (first order) effects have variance $V_1 = 2^{-k}(1+\rho)^{k-1}(1-\rho) = \mu V$, while the second order interactions have variance $V_2 = \mu^2 V$, $\mu = (1-\rho)/(1+\rho)$.

Suppose $k = 5$ and that the fractional factorial is to be generated by the relation $I = \{A, B, C\} = \{C, D, E\}$, but with the freedom to determine the assignment of numbers 1-5 to letters A-E. In particular, this design is the only (within factor relabelings) 2^{5-2} fractional factorial of resolution III. Let $\boldsymbol{m}_1 = (-1, -1, 1, 1, 1)$, $\boldsymbol{m}_2 = (1, 1, 1, -1, -1)$, $\boldsymbol{m}_3 = (-1, 1, -1, 1, -1)$ be the basic runs of (3.2). The design matrix (with runs across rows and rows properly ordered),

the first column of the covariance matrix $\boldsymbol{R}$, the matrix $\boldsymbol{E}$ of (3.4), and the corresponding eigenvalues (up to a factor of 32) are displayed in Table 1 (we use $\pm$ rather than constantly adjoining a 1).

Table 1

Design (A-E)	$\boldsymbol{R}(1,\cdot)$	$\boldsymbol{E}$	diag (Λ)
$+\ +\ +\ +\ +$	1	$+\ +\ +\ +\ +\ +\ +\ +$	$(1+\rho^2)^2+4\rho^3$
$-\ -\ +\ +\ +$	ρ^2	$+\ -\ +\ -\ +\ -\ +\ -$	$1-\rho^4$
$+\ +\ +\ -\ -$	ρ^2	$+\ +\ -\ -\ +\ +\ -\ -$	$1-\rho^4$
$-\ -\ +\ -\ -$	ρ^4	$+\ -\ -\ +\ +\ -\ -\ +$	$(1-\rho^2)^2$
$-\ +\ -\ +\ -$	ρ^3	$+\ +\ +\ +\ -\ -\ -\ -$	$(1+\rho^2)^2-4\rho^3$
$+\ -\ -\ +\ -$	ρ^3	$+\ -\ +\ -\ -\ +\ -\ +$	$1-\rho^4$
$-\ +\ -\ -\ +$	ρ^3	$+\ +\ -\ -\ -\ -\ +\ +$	$1-\rho^4$
$+\ -\ -\ -\ +$	ρ^3	$+\ -\ -\ +\ -\ +\ +\ -$	$(1-\rho^2)^2$

Consider the mean effect and the five main effects: A maps to 5 (the levels down the first column are found in the sixth column of $\boldsymbol{E}$), B maps to 1, C to 4, D to 2 and E to 6, the 0th order effect always maps to 0. From Proposition 3.3 one finds that posterior covariances are zero while the posterior variances are as follows:

$\mathrm{Var}(\Gamma_0|X(\boldsymbol{t}),\ \boldsymbol{t}\text{ in }F) = V - V^2/[(1+\rho^2)^2+4\rho^3]$,
$\mathrm{Var}(\Gamma_{\{i\}}|X(\boldsymbol{t}),\ \boldsymbol{t}\text{ in }F) = \mu V - \mu^2V^2/(1-\rho^4)$ if i is assigned to A, B, D or E,
$\mathrm{Var}(\Gamma_{\{i\}}|X(\boldsymbol{t}),\ \boldsymbol{t}\text{ in }F) = \mu V - \mu^2V^2/[(1+\rho^2)^2-4\rho^3]$ if i is assigned to C.

Since $\lambda_4 > \lambda_1 = \lambda_2 = \lambda_5 = \lambda_6$, the final posterior variancer is larger than the four preceding it and any one factor thought to be of secondary interest would be assigned to this label. However if one sets up criteria such as minimizing the generalized variance of the Γ's or minimizing the largest of the posterior variances (perhaps weighted according to order), it is clear that factor assignment is unimportant.

Going further with the same situation, suppose that we have equal interest in the main effects but that $\Gamma_{\{1,2\}}$ also warrants attention. Without loss we can consider the three cases of assigment: factor 1 to label A and factor 2 to label B, C or D. In the first of these cases, $\{1,2\}$ and C agree with the 5th column of $\boldsymbol{E}$ (equivalently, map to the integer 4 in our notation); in the second, $\{1,2\}$ and B map to 1 (the second column of $\boldsymbol{E}$); in the third, $\{1,2\}$ (alone) maps to 7. Since λ_7 corresponds to the smallest of the distinct eigenvalues for any ρ in $(0,1)$, an assignment of factor 2 to label D (equivalently E) results in the smallest

posterior variance of $\Gamma_{\{1,2\}}$ while retaining the same set of remaining posterior variances. The way we have structured things, the least important of the factors 3, 4 and 5 could then be assigned to C.

Retaining the context, one could formally consider D-optimality with respect to $\Gamma_{\{1\}}$, $\Gamma_{\{2\}}$, $\Gamma_{\{3\}}$, $\Gamma_{\{4\}}$, $\Gamma_{\{5\}}$ and $\Gamma_{\{1,2\}}$, (unnormalized by interaction order, say) of the various assignments. This reduces to another comparison of the choices: factors 1 to A and 2 to B; factors 1 to A and 2 to C; factors 1 to A and 2 to D. In the first case $\{1,2\}$ is aliased with the factor, say 3, assigned to C; in the second of the $\{1,2\}$ is aliased with the factor assigned to B. Calculation of determinants then shows that the last of these assignments is preferred to the second when $\rho > .1552$, and it is preferred to the first if $\rho > .1086$. Thus minimizing the posterior variance of $\Gamma_{\{1,2\}}$ also gives the D-optimum design, but only when ρ is large enough.

Bringing in a "less interesting" factor might also be accomplished at the prior specification stage. Consider assigning labels to the 2^{5-2} design of Table 1 when the fifth factor plays the special role, and keep open the possibility that the interaction between factors 1 and 2 might also be emphasized. Let a non-isotropic correlation be given by $R(\boldsymbol{s},\boldsymbol{t}) = (\theta/\rho)^{(1/2)|s_5-t_5|}\rho^{d(s,t)}$, where $\theta > \rho$. Here the correlation, θ, between sites which differ only in the fifth factor is higher than that ρ, gotten by changing only the jth, $j \neq 5$. If the fifth factor is labelled C, the first column of $\boldsymbol{R}$ and the column of the λ's corresponding to those in Table 1 are given in Table 2a; if the fifth factor is labelled E, the same entries are in Table 2b (note that if the fifth factor is labelled A, B or D, one merely permutes the λ's in Table 2b).

Table 2

2a: 5 as C		2b: 5 as E	
$\boldsymbol{R}(1,\cdot)$	diag (Λ)	$\boldsymbol{R}(1,\cdot)$	diag (Λ)
1	$(1+\rho^2)^2+4\theta\rho^2$	1	$(1+\theta\rho)(1+\rho^2)+2\rho^2(\theta+\rho)$
ρ^2	$1-\rho^4$	ρ^2	$(1+\theta\rho)(1-\rho^2)$
ρ^2	$1-\rho^4$	$\theta\rho$	$(1-\theta\rho)(1+\rho^2)+2\rho^2(\theta-\rho)$
ρ^4	$(1-\rho^2)^2$	$\theta\rho^2$	$(1-\theta\rho)(1-\rho^2)$
$\theta\rho^2$	$(1+\rho^2)^2-4\theta\rho^2$	$\theta\rho^2$	$(1+\theta\rho)(1+\rho^2)-2\rho^2(\theta+\rho)$
$\theta\rho^2$	$1-\rho^4$	$\theta\rho^2$	$(1+\theta\rho)(1-\rho^2)$
$\theta\rho^2$	$1-\rho^4$	ρ^3	$(1-\theta\rho)(1+\rho^2)-2\rho^2(\theta-\rho)$
$\theta\rho^2$	$(1-\rho^2)^2$	ρ^3	$(1-\theta\rho)(1-\rho^2)$

For D-optimality (minimizing the generalized variance) over the main effects together with the $\{1,2\}$ interaction, we consider the following labellings: (A, B,

C, D, E) goes on $(1,2,5,3,4)$, $(1,3,5,2,4)$, $(1,2,3,4,5)$, $(1,3,2,4,5)$, $(3,4,1,2,5)$ and $(1,3,4,2,5)$. The aliasings are, by turn, $\{1,2\}$ with $\{5\}$, none, $\{1,2\}$ with $\{3\}$, $\{1,2\}$ with $\{3\}$, $\{1,2\}$ with $\{5\}$, none. For the first two labellings one refers to Table 2a for the eigenvalues, for the last four one looks to Table 2b. Finally one computes the variance terms and, if necessary, the covariance terms from (2.2) and Proposition 3.3.

The relevant covariance matrices are accomodating enough to be diagonal when there is no aliasing and diagonal except for one 2×2 submatrix when aliasing is present. Moreover when the comparison reduces to that of 2×2 subdeterminants, the algebra is simple. In particular, the second labelling is preferred to the first provided $\mathrm{Var}(\Gamma_{\{5\}}) = V_{\{5\}} < \lambda_4 - \lambda_7$. In the present circumstance this means

$$2^{-5}(1+\rho)^4(1-\theta) < (1+\rho^2)^2 - 4\theta\rho^2 - (1-\rho^2)^2.$$

Thus $(1+\rho)^4 < 128\rho^2$, or $\rho > .1086$. There is no dependence on θ and this comes to the same comparison as was made earlier.

Next compare the third, fourth and fifth labelling (involving aliasing) with the last (which does not). The same type of calculations show that (1, 3, 4, 2, 5) is preferred over

$(1,2,3,4,5)$ if $\mathrm{Var}\,(\Gamma_{\{3\}}) = V_{\{3\}} = 2^{-5}(1+\rho)^3(1-\rho)(1+\theta) < \lambda_4 - \lambda_7$;

$(1,3,2,4,5)$ if $\mathrm{Var}\,(\Gamma_{\{3\}}) < \lambda_1 - \lambda_7$;

$(3,4,1,2,5)$ if $\mathrm{Var}\,(\Gamma_{\{5\}}) = V_{\{5\}} = 2^{-5}(1+\rho)^4(1-\theta) < \lambda_6 - \lambda_7$.

Generally speaking these conditions all hold, but not always, and checking them requires more than algebraic manipulation. Accordingly, we have made a grid search over the region $0 < \rho < \theta < 1$ in order to compare all six labellings. The sense of what results is that (1, 3, 5, 2, 4) is best for most parameter values. It is beaten by the choice (1, 2, 5, 3, 4) when ρ is small. Both are beaten by $(3,4,1,2,5)$ when θ is small, say $\theta < .1$, while $(1,3,4,2,5)$ is best when ρ is large, say $\rho > .8$.

Now leave interactions aside and concentrate on the prediction of unobserved site values. The D-optimality problem is one of minimizing the generalized posterior variance of unobserved process values or, equivalently, maximizing the generalized variance of observed values. We take the correlation function from (2.2) with $\rho_1 \le \rho_2 \le \rho_3 \le \rho_4 \le \rho_5$, say, the design from Table 1, and the task of labelling the factors.

From Proposition 3.2, we may content ourselves with maximizing the product of the resulting λ's. Thus in Table 1, the entries of the $\boldsymbol{R}$ column are 1, $\rho_A\rho_B$, $\rho_D\rho_E$, $\rho_A\rho_B\rho_D\rho_E$, etc., and the λ's arise as the inner product of the $\boldsymbol{R}$ column

with the columns of $\boldsymbol{E}$. Given specific choices of ρ's, a computer search can be accomlished over 120 possible assignments. In fact the existence of interchangeable pairs (A, B) and (D, E) reduces consideration to 15 different assignments. For two sets of ρ's, Table 3 lists the best six assignments together with the associated determinant value. The minor altering of ρ values leaves the D-optimum design unchanged, while some differences in the complete orderings do surface.

Table 3

$\rho_i = i/10$				$\rho_i = 2^{i-6}$			
C	(A,B)	(D,E)	det $(\boldsymbol{R})$	C	(A,B)	(D,E)	det $(\boldsymbol{R})$
5	(1,4)	(2,3)	.9606	5	(1,4)	(2,3)	.9983
4	(1,5)	(2,3)	.9576	5	(1,3)	(2,4)	.9979
5	(1,3)	(2,4)	.9545	4	(1,5)	(2,3)	.9976
3	(1,5)	(2,4)	.9494	3	(1,5)	(2,4)	.9970
4	(1,3)	(2,5)	.9419	5	(1,2)	(3,4)	.9957
3	(1,4)	(2,5)	.9400	4	(1,3)	(2,5)	.9950

G-optimal designs are computationally feasible at this level of complexity as well, but we defer a more systematic examination of this and other types of optimality of specific fractional factorials to a subsequent paper.

Acknowledgments

The research of Mitchell and Morris was sponsored by the Applied Mathematical Sciences Research Program, Office of Energy Research, U. S. Department of Energy under contract DE-AC05-84OR21400 with the Martin Marietta Energy System, Inc. Ylvisaker's research was supported in part by NSF Grant DMS 92-13264. The authors thank Ed D'Azevedo, Persi Diaconis, Mark Johnson and Chuck Romine for helpful conversations in the course of this research. Two referees have been generous with time and useful comments.

Toby Mitchell died on June 5, 1993 of complications of leukemia. His coauthors have lost a close friend, a marvelous colleague and an especially thoughtful voice, as have many others.

References

Currin, C., Mitchell, T. J., Morris, M. and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions with applications to the design and analysis of computer experiments. J. Amer. Statist. Assoc. **86**, 953-963.

Easterling, R. G. (1989). Comment on "Design and analysis of computer experiments", by J. Sacks, W. J. Welch, T. J. Mitchell and H. P. Wynn. Statist. Sci. **4**, 425-427.

John, P. W. M. (1971). Statistical Design and Analysis of Experiments. Macmillan, New York.

Johnson, M. E., Moore, L. M. and Ylvisaker, D. (1990). Minimax and maximin distance designs. J. Statist. Plann. Inference **26**, 131-148.

Kimeldorf, G. S. and Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Statist. **41**, 495-502.

Letac, G. (1981). Problemes classiques de probabilite sur un couple de Gel'fand. Analytical Methods in Probability Theory. Lecture Notes in Math. **861**, 93-120. Springer, Berlin.

MacWilliams, F. J. and Sloane, N. J. A. (1977). The Theory of Error Correcting Codes. North Holland, Amsterdam.

McKay, M. D., Conover, W. J. and Beckman, R. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **21**, 239-245.

Morris, M. D. and Mitchell, T. J. (1995). Exploratory designs for computational experiments. J. Statist. Plann. Inference, in press.

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

Plackett, R. L. and Burman, J. P. (1943). The design of optimum multifactorial experiments. Biometrika **33**, 305-325.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments (with comments). Statist. Sci. **4**, 409-435.

Tang, B. (1993). Orthogonal array-based Latin hypercubes. J. Amer. Statist. Assoc. **88**, 1392-1397.

Yaglom, A. M. (1961). Second order homogeneous random fields. 4th Berkeley Symp. II, 593-622.

Ylvisaker, D. (1987). Prediction and design. Ann. Statist. **15**, 1-19.

Mathematical Sciences Section, EPM Division, Oak Ridge National Laboratory, Building 6012, P.O. Box 2008, Oak Ridge, TE 37831-6367, U.S.A.

Department of Mathematics, University of California, Los Angeles, CA 90024, U.S.A.

(Received December 1993; accepted February 1995)