Statistica Sinica **11**(2001), 605-630

BAYESIAN OPTIMAL FRACTIONAL FACTORIALS

M. Kathleen Kerr

The Jackson Laboratory

Abstract: We take a Bayesian approach to choosing among 2^{k-p} fractional factorials. Experimental observations are thought of as realizations of a stationary Gaussian process X operating on the design space. Pre-experimental knowledge is formally incorporated in the distribution of X. Rather than demanding a precise prior distribution for X, we seek designs that are optimal for families of priors, making the results robust. We examine Bayesian D-, A-, G-, E-, and **c**-optimality, paying closest attention to D-optimality. Within a family of processes, we characterize D-optimal designs for nearly-independent and nearly-dependent priors. Often the maximum resolution–minimum aberration design is found optimal in all cases. However, for some k and p, a second design turns out to be optimal for certain subfamilies of processes.

Key words and phrases: Asymptotic D-optimality, Bayesian prediction, fractional factorials, Gaussian processes, maximin distance designs, resolution, stationary processes, two-level factors, word length pattern.

1. Introduction

Fractional factorials are the classical designs for experiments with k two-level factors. For fixed k and fraction size, an experimenter must choose among many candidate fractional factorials. We take a Bayesian approach, incorporating prior understanding about the relationships between experimental observations into our choice of design. Specifically, we think of experimental observations as realizations of a stochastic process X operating on the design space. We summarize pre-experimental knowledge in the distribution of X, which we assume is stationary and Gaussian. The stationarity assumption is meant to ensure a kind of "impartiality" for the distribution of X. Design criteria can be formulated in terms of the distribution of X conditioned on observations in a design.

Our approach differs from previous Bayesian design work that discusses estimation of parameters for a fixed model. For example, Chaloner (1984) studies optimal Bayesian designs for linear models, assuming a proper normal prior distribution for the parameters. DuMouchel and Jones (1994) propose modifying D-optimal designs to incorporate model uncertainty. In their explicitly model-oriented approach, Meyer, Steinberg and Box (1996) propose Bayesian methods for designing follow-up experiments when confounding leaves more than one

model consistent with the data after an initial experiment. They start with prior probabilities on models rather than on parameters.

Mitchell, Morris and Ylvisaker (1995) study stationary Gaussian processes for the purpose of choosing a fractional factorial. Here we generalize their formulation to include experiments with observational error instead of only considering deterministic settings such as computer experiments. Other closely related work is that of Toman (1994), who considers designs for multiple two- and three-level factors by imagining observations as realizations of a stochastic process. The processes considered there assume exchangeability and are a special case of those we consider. Toman's design criteria relate to minimizing the posterior variances of linear functions of responses. Classical fractional factorials tend to be optimal under these criteria.

In addressing design questions we are motivated by the problem of choosing among fractional factorials. We do not take up the important problem of finding them. References for this topic include Chen (1998), Chen, Sun and Wu (1993), Chen and Wu (1991), Franklin (1985), Laycock and Rowley (1995) and Tang and Wu (1996).

Section 2 establishes notation, develops the Bayesian framework, and identifies families of stationary processes that should be most appropriate for design. Section 3 discusses fractional factorials and the distribution of stationary processes conditioned on these designs. Using those results, Section 4 formulates design criteria. We apply these criteria to the relatively simple case of half-fractions in Section 5. We then concentrate on D-optimality. Sections 6 and 7 employ asymptotics to simplify the D-optimality criterion. In most cases it turns out to be straightforward to identify asymptotically D-optimal fractions. Section 8 discusses the asymptotic criteria, giving examples of optimal designs for particular values of k and p.

2. Framework

2.1. The design space T, words, and interactions

Use "1" and "-1" to denote the two levels for each of k binary factors in an experiment. The design space is $T = \{-1, 1\}^k$, all k-dimensional vectors with entries "1" and "-1". The elements of T are the possible experimental *runs*.

The space T forms a group via component-wise direct multiplication with identity $\mathbf{1}$ and every element self-inverse. In addition, T is a vector space over $\mathbb{F}_2$. Although we generally prefer multiplicative notation, as a vector space it is more usual to represent T by mapping "1" to "0" and "-1" to "1". Then binary vector addition replaces component-wise multiplication. T is also a metric space under Hamming distance d, where $d(\mathbf{t}, \mathbf{s})$ is the number of components in which the runs $\mathbf{t}$ and $\mathbf{s}$ disagree. Let $|\mathbf{t}| \equiv d(\mathbf{1}, \mathbf{t})$.

Use $1, \ldots, k$ to represent the k factors in the experiment. Subsets of $\{1, \ldots, k\}$ are *words* and the collection of words is $\mathcal{W}$. A word $W \in \mathcal{W}$ has *length* $|W|$. The words $\mathcal{W}$ form a group, where the product of two words W and U is their disjoint union, $(W \cup U)/(W \cap U)$. The empty set is the identity and every word is self-inverse. An alternate representation of words is as k-dimensional vectors, indexed by the k factors, with a "1" for all factors in W and "0" otherwise. The disjoint union of words corresponds to binary addition of vectors. This notation makes it obvious that $\mathcal{W}$ and T are isomorphic groups. Like T, $\mathcal{W}$ forms a vector space over $\mathbb{F}_2$.

Every word is also a function on $T \to \{-1, 1\}$, defined by $W(\mathbf{t}) = \prod_{i \in W} t_i$. For any real-valued function x on T, the *interaction term associated with a word* W is $\Gamma_W(x) = 2^{-k} \sum_{\mathbf{t}} W(\mathbf{t}) x(\mathbf{t})$. The function x can subsequently be recovered from the $\Gamma' s$ as $x(\mathbf{t}) = \sum_W W(\mathbf{t}) \Gamma_W(x)$. Our interest lies in interactions for random functions.

2.2. Stationary processes on T

In our notation, X is a mean zero, Gaussian process on T available for observation (possibly with error). Our primary assumption is that X is a stationary process, meaning that for any $\mathbf{u} \in T$, $\{Z(\mathbf{t})\}$ is distributed as $\{X(\mathbf{t})\}$, where $Z(\mathbf{t}) = X(\mathbf{ut})$. In words, stationarity means the correlation between $X(\mathbf{t})$ and $X(\mathbf{s})$ depends only on the factors where $\mathbf{t}$ and $\mathbf{s}$ differ. For $\mathbf{t} \in T$, define $r(\mathbf{t}) = E[X(\mathbf{1})X(\mathbf{t})]$, so that $E[X(\mathbf{s})X(\mathbf{t})] = r(\mathbf{st})$.

In an experiment on k binary factors, we suppose the ith observation at $\mathbf{t} \in T$ is $X_i(\mathbf{t})$, where

$$X_i(\mathbf{t}) = X(\mathbf{t}) + \epsilon_i(\mathbf{t}). \tag{1}$$

The error process $\epsilon_i(\mathbf{t})$ is assumed to be a mean-zero Gaussian process. We further assume the error process is independent of X, and $E[\epsilon_i(\mathbf{t})\epsilon_j(\mathbf{s})] = \sigma^2 \delta_{ij} \delta_{\mathbf{ts}}$. Notice that independence among $\epsilon(\mathbf{t})$ is a special case of stationarity. Since the sum of stationary processes is again stationary, for fixed i, X_i in (1) is stationary. For the case of error-free experimentation (such as computer experiments), set $\sigma^2 = 0$.

A process X on T is a random function on T, so X has interactions as defined in Section 2.1. A key fact is that a Gaussian process X is stationary if and only if its interactions $\{\Gamma_W(X)\}$ are independent (Mitchell, Morris and Ylvisaker (1995), Proposition 2.1). Then, since the interactions have mean zero, their joint distribution is completely specified by their variances. Let $\mathrm{v}_W = \mathrm{Var}\,(\Gamma_W(X))$. Then $\mathrm{v}_W = 2^{-k} \sum_{\mathbf{t}} W(\mathbf{t}) r(\mathbf{t})$ gives the interaction variances in terms of the covariances of the process. Similarly, covariances of the process can be recovered from interaction variances as $r(\mathbf{t}) = \sum_W W(\mathbf{t}) \mathrm{v}_W$.

Example 1. (Mitchell, Morris and Ylvisaker (1995)). Start with positive correlations $\rho_1, \ldots, \rho_k$ and set $r(\mathbf{t}) = \prod_{i:t_i \neq -1} \rho_i$, which defines a stationary process by $E[X(\mathbf{s})X(\mathbf{t})] = r(\mathbf{st})$. The process has interaction variances $\mathrm{v}_W = 2^{-k} \prod_{i \in W}(1-\rho_i) \prod_{i \notin W}(1+\rho_i)$. When $\rho_i \equiv \rho$, $r(\mathbf{t}) = \rho^{|\mathbf{t}|}$ and $\mathrm{v}_W = 2^{-k}(1-\rho)^{|W|}(1+\rho)^{k-|W|}$.

Use R to denote the covariance matrix of X on T and $\mathbf{X}$ for the 2^k vector of $X(\mathbf{t})$. For a stationary process X, the random vector $\mathbf{X}$ is transformed to 2^k *independent* interactions Γ_W. We summarize this transformation using matrix notation. Write $\mathbf{\Gamma}$ for the vector of 2^k interactions and $O = O_k$ for the $2^k \times 2^k$ matrix that transforms $\mathbf{X}$ to $\mathbf{\Gamma}$. Then $\mathbf{\Gamma} = O\mathbf{X}$. Next consider the covariance matrix R for $\mathbf{X}$. Since $\mathbf{\Gamma} = O\mathbf{X}$, $V = \mathrm{Var}\,(\Gamma) = ORO'$. Because the interactions for a stationary process are independent, V is a diagonal matrix containing the interaction variances v_W. Moreover, one can check from the definition of the interactions as contrasts of the $\{X(\mathbf{t})\}$ that O is an orthogonal matrix. This is true because (i) every homomorphism in the word group except $\emptyset$ is "1" on exactly half the elements of T and "-1" on the other half, and (ii) any two word homomorphisms agree on exactly half the elements of T.

Since O is orthogonal and $2^{k/2}O$ is orthonormal, we can rewrite $V = ORO'$ as $R = (2^{k/2}O')2^kV(2^{k/2}O)$ and obtain a diagonalization of R. Since R is symmetric, 2^kV gives its eigenvalues. In other words, the eigenvalues of R are nothing more than the interaction variances for the process, multiplied by 2^k.

2.3. Families of priors

There are families of stationary processes on T that are particularly appropriate for design. Three such families are *isotropic*, *orderly*, and *tame* processes.

An *isotropic* process is a stationary process for which $r(\mathbf{t})$ depends on $|\mathbf{t}|$ alone. It turns out that X is isotropic if and only if v_W depends only on $|W|$ (Mitchell, Morris and Ylvisaker (1995), Proposition 2.2). A stationary process X on T has 2^k parameters. The distribution of X is specified by 2^k covariances $r(\mathbf{t})$ or, alternatively, by 2^k interaction variances v_W. For an isotropic process, $r(\mathbf{t}) = r(|\mathbf{t}|)$ and $\mathrm{v}_W = \mathrm{v}_{|W|}$, reducing the number of parameters to $k+1$. This is a substantial reduction for large k. For an isotropic process, let $\mathbf{r}$ denote the vector of covariances $(r_0, \ldots, r_k)'$ and $\mathbf{v}$ denote the vector of interaction variances $(\mathrm{v}_0, \ldots, \mathrm{v}_k)'$.

In analyzing the data from a multifactor experiment, researchers standardly assume higher order effects are more likely to be negligible. Incorporating this assumption into a stationary process X on T, the distribution of X should assign less variability to higher order interactions. Processes that incorporate this assumption have the property that $W \subset U \Longrightarrow \mathrm{v}_W > \mathrm{v}_U$, referred to as *nested decreasing* interaction variances. In addition, one naturally prefers processes that

have positive covariances with the property that $R(\mathbf{t},\mathbf{s}) > R(\mathbf{t},\mathbf{u})$ whenever $\mathbf{s}$ is *between* $\mathbf{t}$ and $\mathbf{u}$, i.e., whenever $d(\mathbf{t},\mathbf{u}) = d(\mathbf{t},\mathbf{s}) + d(\mathbf{s},\mathbf{u})$. This is really the same kind of partial ordering on covariances as the partial ordering of nested decreasing interaction variances, and so is also referred to as nested decreasing. We call stationary processes with both of these properties *orderly* processes.

For an isotropic process, the property of nested decreasing interaction variances reduces to $\mathrm{v}_{i-1} > \mathrm{v}_i$ and nested decreasing covariances means $r_{i-1} > r_i$ for $i = 1,\ldots,k$. It is not enough to assume the v_i decrease to guarantee the r_i decrease. Assuming interaction variances are convex and decreasing also does not suffice for all k. However, a stronger assumption on the v_i yields a stronger result for the r_i.

As in Mitchell, Morris and Ylvisaker (1995), define difference operators D^m for a finite sequence a_j, $j = 0,\ldots,k$, as follows:

$$D^m a_j = \sum_{h=0}^{m} (-1)^h \binom{m}{h} a_{j+h}, \quad \text{for } j+m \le k.$$

Call the finite sequence a_j, $j = 0,\ldots,k$, *completely monotone* provided $D^m a_j \ge 0$ for all $j+m \le k$. Kerr ((1999), Proposition 3.3.1) shows that an isotropic process has completely monotone correlations r_i if and only if it has completely monotone interaction variances v_j. This result leads naturally to defining the class of *tame* processes on T — isotropic processes with completely monotone interaction variances and completely monotone correlations. Tame processes behave nicely, as they are "balanced" across factors (isotropic) and their correlations and interaction variances decrease "smoothly" (completely monotone). The set of tame processes is in fact smaller than the set of orderly isotropic processes. We picture the families of processes as in Figure 1.

Example 1.(continued) The isotropic process with $r(\mathbf{t}) = \rho^{|\mathbf{t}|}$ and $\mathrm{v}_i = 2^{-k}(1+\rho)^{k-i}(1-\rho)^i$, $0 < \rho < 1$, is a tame process.

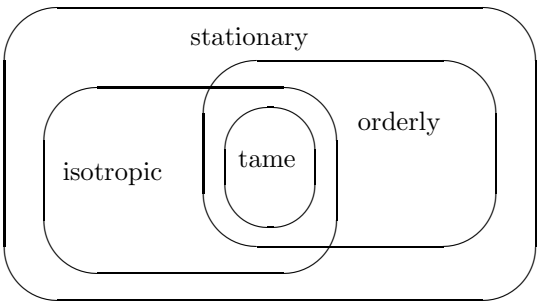

Figure 1. Families of processes on T.

3. Designs and Conditioning

A regular *fractional factorial* design is specified by p independent words and the "defining relation" $\emptyset = W_1 = \cdots = W_p$. The set of runs in the design is then $F = \{\mathbf{t} \in T : W_i(\mathbf{t}) = 1,\ i = 1, \ldots, p\}$, and $|F| = 2^{k-p}$. Note F is a subgroup of T. The words $W_1, \ldots, W_p$ specifying a fractional factorial F span a subgroup of $\mathcal{W}$, which we call $\mathcal{A}_F$. The subgroup $\mathcal{A}_F$ contains the *defining words*. Call cosets of $\mathcal{A}_F$ in $\mathcal{W}$ *alias sets* and denote a generic alias set as $\mathcal{A}$. There are 2^{k-p} alias sets, each with 2^p words, partitioning the word group. Note this partition of words induces a partition of interactions, so one can naturally speak of alias sets of interactions without confusion.

A few more bits of notation are helpful. For a fractional factorial F define $D_i = |\{\mathbf{f} \in F : |\mathbf{f}| = i\}|$ and $L_i = |\{W \in \mathcal{A}_F : |W| = i\}|$. The vectors $(D_0, \ldots, D_k)'$ and $(L_0, \ldots, L_k)'$ are the *design distance* and *word length* vectors, respectively. Of course, D_0 and L_0 are always 1. The smallest index $i > 0$ such that $L_i > 0$ is commonly known as the *resolution* of F. If F has resolution Res then L_{Res} is the *aberration* of F. The most common criterion for choosing fractional factorials in the literature is the maximum resolution–minimum aberration criterion of Fries and Hunter (1980).

3.1. Stationary processes and fractional factorials

From Section 2.2, the covariance matrix R for X can be written $R = (2^{k/2}O')2^kV(2^{k/2}O)$, where $2^{k/2}O$ is orthonormal and V is a diagonal matrix of interaction variances. In total, the interactions are an orthogonal transform of a stationary process that gives the eigenvalues for its covariance matrix. It turns out that the structure of R is "repeated" in certain principal submatrices of R, namely submatrices corresponding to X restricted to a subgroup of T.

The definition of stationarity is essentially invariance under group translation of runs. Moreover, subgroups of T are closed under multiplication. Consider X as a process on a subgroup F of T. For any given $\mathbf{u} \in F$, $Z(\mathbf{f}) = X(\mathbf{uf})$ is a well-defined process on F because $\mathbf{uf} \in F$. Since X is stationary on T, $\{Z(\mathbf{t})\}$ is distributed as $\{X(\mathbf{t})\}$. In other words, a stationary process restricted to a subgroup is also a stationary process on the subgroup. Let R_F denote the covariance matrix of X restricted to F. Mitchell, Morris and Ylvisaker ((1995), Proposition 3.2) show how to diagonalize R_F and give its eigenvalues as proportional to linear combinations of the covariances $r(\mathbf{t})$. The alternate derivation in Kerr ((1999), Section 5.1) uses the group structure of T and gives the eigenvalues as simple sums of the interaction variances v_W. This is described next.

For each alias set $\mathcal{A}$, define the *alias interaction* $\Gamma_{\mathcal{A}} = \sum_{W \in \mathcal{A}} \Gamma_W$. Since alias sets are disjoint, alias interactions inherit independence from the Γ_W. It turns

out that the $\Gamma_{\mathcal{A}}$ can be written solely as linear combinations of $X(\mathbf{f})$ for $\mathbf{f} \in F$. In matrix notation, let $\mathbf{\Gamma}_{\mathcal{A}}$ be the vector of the 2^{k-p} alias interactions, $\mathbf{X}_F$ be the vector of $X(\mathbf{f})$ for $\mathbf{f} \in F$, O_F be the $2^{k-p} \times 2^{k-p}$ matrix that transforms $\mathbf{X}_F$ to $\mathbf{\Gamma}_{\mathcal{A}}$, and R_F be the covariance matrix for X restricted to F. It is shown that $\mathbf{\Gamma}_{\mathcal{A}} = O_F \mathbf{X}_F$, and so $V_F = \mathrm{Var}\,(\mathbf{\Gamma}_{\mathcal{A}}) = O_F R_F O_F'$. Since the alias interactions are independent, V_F is diagonal with the variances of the alias interactions along the diagonal. Letting $\mathrm{v}_{\mathcal{A}} = \mathrm{Var}\,(\Gamma_{\mathcal{A}})$, we have $\mathrm{v}_{\mathcal{A}} = \sum_{W \in \mathcal{A}} \mathrm{v}_W$.

Just as O was an orthogonal transformation of X on T, O_F is an orthogonal transformation of X on F. Since $2^{(k-p)/2} O_F$ is orthonormal, write $R_F = (2^{(k-p)/2} O_F')2^{k-p} V_F (2^{(k-p)/2} O_F)$. As noted, V_F is diagonal, so $2^{k-p} V_F$ gives the eigenvalues of R_F. An eigenvalue of R_F is 2^{k-p} times the sum of the interaction variances in an alias set.

3.2. Conditional distributions

In our Bayesian framework, design criteria will be formulated in terms of the conditioned process $\{X(\mathbf{t}) | X(\mathbf{f}) + \epsilon(\mathbf{f}), \mathbf{f} \in F\}$. Mitchell, Morris and Ylvisaker (1995) give the distribution of X conditioned on error-free observations on a fractional factorial. Kerr ((1999), Section 5.2) generalizes those results to observation with error and also gives the parameters of the posterior distribution in terms of the interaction variances of the unconditioned process.

For notational ease, Var_F and Cov_F denote posterior variances and covariances given observations $X(\mathbf{f}) + \epsilon(\mathbf{f})$ for $\mathbf{f} \in F$, and σ_p^2 denotes $2^{-(k-p)}\sigma^2$. First, the posterior variances and covariances of the interactions given observation on F are

$$\mathrm{Var}_F(\Gamma_W) = \mathrm{v}_W - \frac{\mathrm{v}_W^2}{\mathrm{v}_{(W\mathcal{A}_F)} + \sigma_p^2}, \tag{2}$$

$$\mathrm{Cov}_F(\Gamma_W, \Gamma_U) = -\frac{\mathrm{v}_W \mathrm{v}_U}{\mathrm{v}_{(W\mathcal{A}_F)} + \sigma_p^2} \delta_{(W\mathcal{A}_F)(U\mathcal{A}_F)}. \tag{3}$$

Note $W\mathcal{A}_F$ is the alias set containing W. Further note that interactions in different alias sets remain independent for the conditioned process.

The posterior variance of $X(\mathbf{t})$ is gotten by calculating $\sum_{\mathcal{A}} \mathrm{Var}_F(\sum_{W \in \mathcal{A}} W(\mathbf{t})\Gamma_W)$:

$$\mathrm{Var}_F(X(\mathbf{t})) = \sum_{\mathcal{A}} \Big[\frac{\mathrm{v}_{\mathcal{A}} \sigma_p^2}{\mathrm{v}_{\mathcal{A}} + \sigma_p^2} + \frac{4}{\mathrm{v}_{\mathcal{A}} + \sigma_p^2} \sum_{\substack{W,U \in \mathcal{A}: \\ W(\mathbf{t}) \neq U(\mathbf{t})}} \mathrm{v}_W \mathrm{v}_U \Big]. \tag{4}$$

4. Bayesian Design Criteria

Experimenters may be most interested in interactions because they correspond to factor effects. Mathematically and practically, for some criteria it

makes sense to discuss good designs for inference on interactions as much as for the actual $X(\mathbf{t})$. We therefore formulate some criteria in terms of the posterior distribution of $X(\mathbf{t})$ and some in terms of the posterior distribution of interactions $\Gamma_W(X)$. As before, when observation is error-free, set $\sigma^2 = 0$.

4.1. D-optimality

The D-optimal fractional factorial of a given size is the fractional factorial that minimizes the posterior generalized variance of X. Equivalently, the D-optimal design maximizes $|R_F + \sigma^2 I|$ (Mitchell, Sacks and Ylvisaker (1994)). From Section 3.1, an eigenvalue of R_F is 2^{k-p} times $\mathrm{v}_{\mathcal{A}}$, the sum of the interaction variances in an alias set $\mathcal{A}$. Thus,

$$|R_F + \sigma^2 I| = \prod_{\mathcal{A}} (2^{k-p}\mathrm{v}_{\mathcal{A}} + \sigma^2) = (2^{k-p})^{2^{k-p}} \prod_{\mathcal{A}} (\mathrm{v}_{\mathcal{A}} + \sigma_p^2). \tag{5}$$

When comparing fractions of the same size (constant p), ignore the power of 2 in (5). Thus the D-optimal fractional factorial of a given size partitions the word group to maximize

$$D(F) = \prod_{\mathcal{A}} (\sigma_p^2 + \sum_{W \in \mathcal{A}} \mathrm{v}_W). \tag{6}$$

Orderly processes, defined in Section 2.3, are a family of stationary processes of particular interest. For orderly processes, certain designs can be immediately excluded from consideration for D-optimality. Lemma 4.1 limits the search for D-optimal designs to designs that use all k factors in the defining words. These are generally considered the best designs in other frameworks. For example, a maximum resolution design always uses all k factors in its defining words (Fries and Hunter (1980)). This provides assurance that the Bayesian approach produces reasonable designs. The proof of Lemma 4.1 is deferred to an appendix.

Lemma 4.1. *Suppose X is a stationary process on T with nested decreasing interaction variances. (For example, suppose X is an orderly process.) Let F be a fractional factorial in which some factor is omitted from every defining word. Then there exists a fraction of the same size F^+ such that $D(F^+) > D(F)$ for $\sigma^2 \geq 0$. In particular, F is not D-optimal.*

4.2. A-optimality and E-optimality

The A-optimal fractional factorial of a given size minimizes the average posterior variance of $X(\mathbf{t}), \mathbf{t} \in T$. We can also speak of the A-optimal fraction for interactions, the design that minimizes the average posterior variance of the interactions. The E-optimal fractional factorial minimizes the largest eigenvalue of

the posterior covariance matrix for X on T. Similarly, the E-optimal design for interactions minimizes the largest eigenvalue for the posterior covariance matrix of interactions.

Let V be the diagonal covariance matrix of the interactions $\{\Gamma_W(X)\}$. Since $R = (2^{k/2}O')2^kV(2^{k/2}O)$ and $2^{k/2}O$ is orthonormal, $\text{trace}(R) = \text{trace}(2^kV)$. Moreover, the eigenvalues of R are the same as the eigenvalues of V. The importance of these remarks is that they are valid in reference to posterior distributions as well. For example, $\text{trace}[\text{Var}_F(\mathbf{X})] = \text{trace}[\text{Var}_F(\Gamma)]$. This discussion proves Proposition 4.1.

Proposition 4.1. *A-optimality of X is equivalent to A-optimality of the interactions. E-optimality of X is equivalent to E-optimality of the interactions.*

To formulate the A-optimality criterion, Proposition 4.1 allows us to sum the posterior variances of the interactions rather than the posterior variances of the $X(\mathbf{t})$. Referring to (2), the A-optimal fractional factorial minimizes $\sum_{\mathcal{A}}\sum_{W\in\mathcal{A}}[\text{v}_W - \text{v}_W^2/(\text{v}_{\mathcal{A}} + \sigma_p^2)]$. Since $\sum_{\mathcal{A}}\sum_{W\in\mathcal{A}}\text{v}_A = \sum_W \text{v}_W$ is constant over designs, the A-optimal fraction maximizes

$$A(F) = \sum_{\mathcal{A}} \frac{\sum_{W\in\mathcal{A}}\text{v}_W^2}{\sigma_p^2 + \sum_{W\in\mathcal{A}}\text{v}_W}. \tag{7}$$

4.3. G-optimality

The G-optimal fractional factorial F of a given size is the fraction that minimizes $\max_{\mathbf{t}}\text{Var}_F(X(\mathbf{t}))$. First, refer to the expression for $\text{Var}_F(X(\mathbf{t}))$ at (4). Notice that for all runs in F, the second part of each summand in (4) vanishes since all words in the same alias set agree (as functions) on design sites. Therefore, the largest posterior variance must occur at a run not in the design. Second, start from $\text{Var}_F(X(\mathbf{t})) = \sum_{\mathcal{A}}\text{Var}_F(\sum_{W\in\mathcal{A}}W(\mathbf{t})\Gamma_W) = \sum_{\mathcal{A}}[\sum_{W\in\mathcal{A}}(\text{v}_W - \text{v}_W^2/(\text{v}_{\mathcal{A}} + \sigma_p^2)) - 2\sum_{W\neq U\in\mathcal{A}}W(\mathbf{t})U(\mathbf{t})\text{v}_W\text{v}_U/(v_{\mathcal{A}} + \sigma_p^2)]$. Since $\sum_{\mathcal{A}}\sum_{W\in\mathcal{A}}\text{v}_W$ is constant over designs, the G-optimal design maximizes

$$\min_{\mathbf{t}\notin F}\sum_{\mathcal{A}}\Big[\frac{\sum_{W\in\mathcal{A}}\text{v}_W^2}{\text{v}_{\mathcal{A}} + \sigma_p^2} + 2\frac{\sum_{W\neq U\in\mathcal{A}}W(\mathbf{t})U(\mathbf{t})\text{v}_W\text{v}_U}{\text{v}_{\mathcal{A}} + \sigma_p^2}\Big]. \tag{8}$$

Note the first part of (8) does not depend on $\mathbf{t}$ and is the criterion for A-optimality at (7).

The fractional factorial that is G-optimal for interactions minimizes

$$\max_W \text{Var}_F(\Gamma_W) = \max_W\Big\{\text{v}_W - \frac{\text{v}_W^2}{\text{v}_{(W\mathcal{A}_F)} + \sigma_p^2}\Big\}. \tag{9}$$

4.4. c-optimality

The **c**-optimal fractional factorial minimizes the posterior variance of $\mathbf{c}'\mathbf{X}$ for some pre-specified vector **c**. Presumably, $\mathbf{c}'\mathbf{X}$ is a linear combination of runs of interest. An example is an experiment where the objective is to predict $\sum_{\mathbf{t}\in T} X(\mathbf{t})$. Then $\mathbf{c} = \mathbf{1}$ and the optimal design minimizes $\mathrm{Var}_F(\sum_{\mathbf{t}\in T} X(\mathbf{t}))$. But $\sum_{\mathbf{t}\in T} X(\mathbf{t})$ is proportional to $\Gamma_\emptyset$, so the **c**-optimal design minimizes $\mathrm{Var}_F(\Gamma_\emptyset) = \mathrm{v}_\emptyset - \mathrm{v}_\emptyset^2/(\sigma_p^2 + \sum_{W\in\mathcal{A}_F} \mathrm{v}_W)$, which means minimizing $\mathrm{v}_{\mathcal{A}_F}$. This criterion can be restated in a more intuitive form. From Kerr ((1999), Proposition 5.1.1), $\mathrm{v}_{\mathcal{A}_F} = 2^{-2(k-p)}\mathrm{Var}\,(\sum_{\mathbf{f}\in F} X(\mathbf{f}))$. Therefore the **c**-optimal fraction minimizes

$$\sum_{\mathbf{f}\in F}\sum_{\mathbf{h}\in F} E[X(\mathbf{f})X(\mathbf{h})] = \sum_{\mathbf{f}\in F}\sum_{\mathbf{h}\in F} E[X(\mathbf{1})X(\mathbf{fh})] = 2^{k-p}\sum_{\mathbf{f}\in F} r(\mathbf{f}). \qquad (10)$$

Examining (10), minimizing $\mathrm{v}_{\mathcal{A}_F}$ is the same as minimizing the average correlation among design points. Informally, to make the best prediction of $\sum_{\mathbf{t}\in T} X(\mathbf{t})$ one should design to maximize the amount of independent information.

5. Optimal Half-Fractions

Half-fractions are well-understood since they are defined by a single, nonempty word. There is little debate that the best design is the maximum resolution design, given by the defining relation $\emptyset = W_k$, where W_k is the unique word of length k. This section shows the answer is the same in our framework. This supports the reasonableness of our Bayesian approach.

Just as with Lemma 4.1, the only property of orderly processes needed for all of the results in this section is that they have nested decreasing interaction variances: $W \subset U \Longrightarrow \mathrm{v}_W > \mathrm{v}_U$. None of the proofs requires the other property of orderly processes that specifies a partial ordering of the covariances. (However, this property comes into play in Section 6.)

5.1. D-, A-, and c-optimality

We first present a lemma that gives a partial ordering of half-fractions with respect to A-optimality. The proof of Lemma 5.1 is deferred to an appendix.

Lemma 5.1. *Suppose X is a stationary process on T with nested decreasing interaction variances that can be observed without error. Let W_l be any word of length l, $2 \le l \le k$, W_{l-1} be a word of length $l-1$, $W_{l-1} \subset W_l$. Let F be the half-fraction defined by $\emptyset = W_l$ and F' be the half-fraction defined by $\emptyset = W_{l-1}$. Then $\sum_{\mathbf{t}} \mathrm{Var}_F(X(\mathbf{t})) < \sum_{\mathbf{t}} \mathrm{Var}_{F'}(X(\mathbf{t}))$.*

Gathering this and results in Section 4, we have the following proposition.

Proposition 5.1. *Suppose X is a stationary process with nested decreasing interaction variances. (For example, suppose X is an orderly process.) Then the*

maximum resolution half-fraction is D-optimal, and A-optimal for both runs and interactions. In addition, it is **c***-optimal for predicting* $\sum_{\mathbf{t}} X(\mathbf{t})$.

Proof. D-optimality follows from Lemma 4.1 and A-optimality follows from Lemma 5.1 and Proposition 4.1. For the **c**-optimality result, recall from Section 4.4 that the best design for predicting $\sum_{\mathbf{t}} X(\mathbf{t})$ minimizes $\sum_{W \in \mathcal{A}_F} \mathrm{v}_W$. Since a half-fraction has only one defining word W other than $\emptyset$, this means choosing W with the smallest interaction variance v_W. This is accomplished by choosing the word of length k.

5.2. Without error: G- and E-optimality

We can improve Proposition 5.1 when X is observed without error. Consider G-optimality for interactions.

Lemma 5.2. *Suppose* X *is a stationary process on* T *with nested decreasing interaction variances and* X *is observed without error* ($\sigma^2 = 0$). *Let* W_l *be any word of length* l, $2 \le l \le k$, W_{l-1} *be a word of length* $l-1$, $W_{l-1} \subset W_l$. *Let* F *be the half-fraction defined by* $\emptyset = W_l$ *and* F' *be the half-fraction defined by* $\emptyset = W_{l-1}$. *Consider any interaction* Γ_W. *Then there exists an interaction* $\Gamma_{W'}$ *such that* $\mathrm{Var}_F(\Gamma_W) < \mathrm{Var}_{F'}(\Gamma_{W'})$.

Proof. Say $W_l = LW_{l-1}$, i.e., W_{l-1} differs from W_l only by the factor L. Referring to (2), since $\sigma^2 = 0$ the posterior variance of an interaction Γ_W for X conditioned on a half-fraction is $(\mathrm{v}_W \mathrm{v}_U)/(\mathrm{v}_W + \mathrm{v}_U)$, where $\{W, U\}$ is an alias set for the half-fraction.

For the first case, consider interactions Γ_W where $W \subseteq W_l$. The design F aliases W and $W_l W$, and exactly one of these words contains L.

There are two sub-cases, $L \notin W$ and $L \in W$.

If W does not contain L, then compare $\mathrm{Var}_F(\Gamma_W)$ to $\mathrm{Var}_{F'}(\Gamma_W) = (\mathrm{v}_W \mathrm{v}_{W_{l-1}W})/(\mathrm{v}_W + \mathrm{v}_{W_{l-1}W})$. Notice

$$\frac{\mathrm{v}_W \mathrm{v}_{W_l W}}{\mathrm{v}_W + \mathrm{v}_{W_l W}} < \frac{\mathrm{v}_W \mathrm{v}_{W_{l-1}W}}{\mathrm{v}_W + \mathrm{v}_{W_{l-1}W}} \iff \mathrm{v}_{W_l W} < \mathrm{v}_{W_{l-1}W}.$$

But $W_l W = LW_{l-1}W \supset W_{l-1}W$, since $L \notin WW_{l-1}$. Since X has nested decreasing interaction variances, $\mathrm{v}_{W_l W} < \mathrm{v}_{W_{l-1}W} \Longrightarrow \mathrm{Var}_F(\Gamma_W) < \mathrm{Var}_{F'}(\Gamma_W)$.

If instead W contains L, compare $\mathrm{Var}_F(\Gamma_W)$ to $\mathrm{Var}_{F'}(\Gamma_{W_l W}) = (\mathrm{v}_{W_l W} \mathrm{v}_{LW})/(\mathrm{v}_{W_l W} + \mathrm{v}_{LW})$. Since $L \in W$, $L \notin LW \Longrightarrow LW \subset W \Longrightarrow \mathrm{v}_{LW} > \mathrm{v}_W \Longrightarrow (\mathrm{v}_W \mathrm{v}_{W_l W})/(\mathrm{v}_W + \mathrm{v}_{W_l W}) < (\mathrm{v}_{W_l W} \mathrm{v}_{LW})/(\mathrm{v}_{W_l W} + \mathrm{v}_{LW})$.

For the second case, consider interactions Γ_W where $W \not\subseteq W_l$. Write $W = W_0 W_1$ with $W_0 \subseteq W_l$, $W_1 \cap W_l = \emptyset$. The argument proceeds exactly as before using W_0 instead of W because adjoining W_1 to every word does not affect any subset or superset relationships.

Next, relate G-optimality for interactions to E-optimality.

Lemma 5.3. *Suppose X is a stationary process on T with nested decreasing interaction variances. When X is observed without error, the half-fraction that is G-optimal for interactions is E-optimal for runs and interactions.*

Proof. For a half-fraction, there are two interaction terms per alias set. Referring to (9) and setting $\sigma^2 = 0$, $\mathrm{Var}_F(\Gamma_W) = \mathrm{Var}_F(\Gamma_U) = -\mathrm{Cov}_F(\Gamma_W, \Gamma_U) = \mathrm{v}_W\mathrm{v}_U/(\mathrm{v}_W + \mathrm{v}_U)$ when F aliases W and U. Written in the appropriate order (aliased interactions adjacent), the posterior covariance matrix of the interaction terms is block-diagonal. The diagonal blocks are 2×2 singular matrices with eigenvalues 0 and $2\mathrm{v}_W\mathrm{v}_U/(\mathrm{v}_W + \mathrm{v}_U)$. Thus the maximum eigenvalue is twice the maximum posterior interaction variance, so G-optimality for interactions implies E-optimality for interactions. Finally, recall from Proposition 4.1 that E-optimality for runs and interactions are equivalent criteria.

Bringing these results together, we have:

Proposition 5.2. *Suppose X is a stationary process with nested decreasing interaction variances observed without error ($\sigma^2 = 0$). Then the maximum resolution half-fraction is G-optimal and E-optimal for both runs and interactions.*

Proof. When X is observed without error, G-optimality for runs is equivalent to A-optimality. This is because the posterior variance for runs in the design is 0, and the posterior variance for unobserved runs is constant. So G-optimality for runs follows from Proposition 5.1. G-optimality for interactions is a corollary of Lemma 5.2, which gives E-optimality by Lemma 5.3.

6. D-Optimality for Near-Independent Orderly Processes

There are an infinite number of distributions for a stationary process X on the k-cube T but only finite number of fractional factorials of a given size. One can think of the space of distributions for X partitioned into subsets such that one design is optimal for every distribution in a subset. Ideally, one would like to know that a single design is optimal for a large class of distributions, such as orderly distributions. A more tractable problem is to identify a few designs that are optimal for subsets of the family of interest and characterize the distributions for which these designs are optimal. In the remainder of this paper we consider only D-optimality and derive results for orderly processes. Because it is both instructive and an important special case, we frequently restate results for orderly isotropic processes.

We approach the problem of identifying D-optimal designs by considering optimal designs for processes near opposite extremes. In this section, a distribution is manipulated so that it converges to complete independence (no correlation

between $X(\mathbf{t})$ and $X(\mathbf{s})$ for $\mathbf{t} \neq \mathbf{s}$). In Section 7 we consider processes near complete dependence (perfect correlation from run to run). Taking limits, we derive criteria for asymptotic D-optimality. The asymptotic criteria derived here and in Section 7 are discussed together in Section 8.

For some fixed k, let $\mathcal{R}$ be the set of possible covariances for a family of processes on T and let X_1, X_2 be independent processes with covariances R_1, $R_2 \in \mathcal{R}$. We say X_i is distributed as R_i, for short. Kerr ((1999), Chapter 4) shows stationary, orderly, isotropic, and tame processes have the following properties:

(P1) Closure under positive linear combinations: $R_1, R_2 \in \mathcal{R}$, $a, b > 0 \Longrightarrow aR_1 + bR_2 \in \mathcal{R}$,
(P2) Closure under direct (element-wise) products: R_1, $R_2 \in \mathcal{R} \Longrightarrow R_1 \otimes R_2 \in \mathcal{R}$,
(P3) Closure under matrix products: R_1, $R_2 \in \mathcal{R} \Longrightarrow R_1 R_2 \in \mathcal{R}$.

These properties ensure that the operations used to manipulate distributions here and in Section 7 do not remove processes from the families to which they belong.

When working with asymptotics, we operate on the distribution of X, not $X + \epsilon$. This keeps the magnitude of error separate from the distribution of X. Without loss of generality, assume the covariance R is a correlation matrix.

6.1. Exponential near-independence

For a correlation R, consider $R^{\otimes n}$, the direct product of R with itself n times. As n gets large, we think of the distribution approaching independence exponentially because $r^n(\mathbf{t})/r^n(\mathbf{s}) \to 0$ whenever $r(\mathbf{t}) < r(\mathbf{s})$. Johnson, Moore and Ylvisaker (1990) study this method of moving a correlation function for an isotropic process toward independence and examine designs in the case of error-free observation. Generalizing their results to stationary processes and observation with error, consider the D-optimality criterion to maximize $|R_F^{\otimes n} + \sigma^2 I|$. The principal terms in the determinant are $(1+\sigma^2)^{|F|} - \sum_{\mathbf{f} \neq \mathbf{f}' \in F} [R(\mathbf{f}, \mathbf{f}')]^{2n}$. For large n the D-optimal fractional factorial minimizes $\sum_{\mathbf{f} \neq \mathbf{f}' \in F} [R(\mathbf{f}, \mathbf{f}')]^{2n}$. As $n \to \infty$ the D-optimal fractional factorial minimizes the maximum correlation among design points.

In the isotropic case, rewrite $\sum_{\mathbf{f} \neq \mathbf{f}' \in F} [R(\mathbf{f}, \mathbf{f}')]^{2n}$ as $\sum_i D_i r_i^{2n}$. For an orderly isotropic process, the asymptotically D-optimal fraction is the maximin distance factorial, i.e., the fraction that maximizes the minimum distance between design points and with the fewest pairs of design points at that distance.

6.2. Linear near-independence

We can bring a correlation R toward independence in a different way, by setting $R^\lambda = \lambda R + (1 - \lambda) I$ and letting $\lambda \to 0$. In this case all correlations

except $r(\mathbf{1})$ approach 0 at the same rate, so we consider this a linear approach to independence. Including the case of error-free observation ($\sigma^2 = 0$), the determinant of the correlation matrix for a design F is $|R_F^\lambda + \sigma^2 I| = |\lambda R_F + (1-\lambda)I + \sigma^2 I| = (1+\sigma^2)^{|F|} - \lambda^2 \sum_{\mathbf{f}\neq\mathbf{f}'\in F}[R(\mathbf{f},\mathbf{f}')]^2 + o(\lambda^2)$. For λ sufficiently close to 0, the D-optimal fractional factorial F minimizes $\sum_{\mathbf{f}\neq\mathbf{f}'\in F}[R(\mathbf{f},\mathbf{f}')]^2$, the average squared correlation among design points. In the isotropic case, the asymptotically D-optimal fraction minimizes

$$\sum_{i=1}^{k} D_i r_i^2. \tag{11}$$

7. D-Optimality for Near-Dependent Orderly Processes

In this section we continue with the asymptotic approach outlined at the beginning of Section 6. Whereas Section 6 found D-optimality criteria for nearly independent processes on T, here we find criteria for nearly dependent processes. We begin with an orderly process and manipulate its distribution so it converges to a degenerate process with perfect correlation from run to run.

7.1. Exponential near-dependence, observation without error

Matrix powers R^n of a correlation matrix R must be normalized to retain "1" along the diagonal. We can calculate a diagonal element of R^n using the decomposition $R = (2^{k/2}O')2^k V(2^{k/2}O)$ where $2^{k/2}O$ is orthonormal. The diagonal matrix V contains the interaction variances v_W. Since $R^n = (2^{k/2}O')2^{nk}V^n(2^{k/2}O)$ and R^n is constant on the diagonal, $R^n(\mathbf{t},\mathbf{t}) = 2^{-k}\mathrm{trace}(R^n) = 2^{nk-k}\mathrm{trace}(V^n) = 2^{nk-k}\sum_W \mathrm{v}_W^n$. Dividing every element of R^n by this factor, we know $2^{k-nk}(\sum_W \mathrm{v}_W^n)^{-1}R^n = 2^{2k}O'((\sum_W \mathrm{v}_W^n)^{-1}V^n)O$ is a correlation matrix.

To omit unimportant constants and simplify presentation, write $R^{(n)}$ for the correlation $2^{k-nk}(\sum_W \mathrm{v}_W^n)^{-1}R^n$ and $\mathrm{v}_W^{(n)}$ for the interaction variances corresponding to $R^{(n)}$. Similarly, $R_F^{(n)}$ means the principal submatrix of $R^{(n)}$ corresponding to the design F. For an orderly process, $\mathrm{v}_\emptyset^{(n)} \to 1$, $\mathrm{v}_W^{(n)} \to 0$ for $W \neq \emptyset$, and $\mathrm{v}_W^{(n)}/\mathrm{v}_U^{(n)} \to 0$ as $n \to \infty$ whenever $W \supset U$.

Letting n get large, we show a process distributed as $R^{(n)}$ tends to a degenerate process on T with complete dependence from run to run. A diagonal element of $(\sum_W \mathrm{v}_W^n)^{-1}V^n$ looks like $\mathrm{v}_U^n/\sum_W \mathrm{v}_W^n$. If the v_W are nested decreasing (in particular, for an orderly process), then $\mathrm{v}_U^n/\sum_W \mathrm{v}_W^n = 1/\sum_W(\mathrm{v}_W/\mathrm{v}_U)^n \to 0$ as $n \to \infty$ for all interaction variances except $\mathrm{v}_\emptyset$. So $R^{(n)} \to 2^{2k}O'V_J O$, where V_J is 0 everywhere except for a 1 in the diagonal position corresponding to $\mathrm{v}_\emptyset$. One can check that $2^{2k}O'V_J O$ is the degenerate correlation matrix J. Because smaller

interaction variances approach 0 faster, we consider this to be an exponential approach towards dependence.

Remark. Note that R has constant row sums and so is proportional to a stochastic matrix. The theory for limits of powers of stochastic matrices can be applied for an alternate proof that $R^{(n)} \to J$.

Return to the question at hand, namely the asymptotically D-optimal fractional factorial for $R^{(n)}$. Different results appear depending on whether or not there is error in observation. We take up the problem for error-free observation here, deferring the case of observation with error to Section 7.2.

The asymptotically D-optimal fractional factorial maximizes $\prod_{\mathcal{A}} \sum_{W \in \mathcal{A}} \mathrm{v}_W^{(n)}$, according to (6), for large n. Now, each alias set $\mathcal{A}$ has a largest interaction variance $\hat{\mathrm{v}}_{\mathcal{A}}$ occurring $n_{\mathcal{A}}$ times. For example, with an orderly process $\hat{\mathrm{v}}_{\mathcal{A}_F} = \mathrm{v}_\emptyset$ and $n_{\mathcal{A}_F} = 1$. Then as $n \to \infty$,

$$\prod_{\mathcal{A}} \sum_{W \in \mathcal{A}} \mathrm{v}_W^{(n)} \sim \prod_{\mathcal{A}} n_{\mathcal{A}} \hat{\mathrm{v}}_{\mathcal{A}}^{(n)}. \tag{12}$$

It is easiest to understand (12) with an example.

Example 2. Consider 2^{6-2} fractional factorials and an orderly isotropic process. The maximum resolution fraction is F_1, defined by $\emptyset = 1234 = 3456 = 1256$, and

$$\begin{aligned} D(F_1) &= (\mathrm{v}_0^{(n)} + 3\mathrm{v}_4^{(n)})(4\mathrm{v}_3^{(n)})^2(2\mathrm{v}_2^{(n)} + 2\mathrm{v}_4^{(n)})^6(\mathrm{v}_1^{(n)} + 2\mathrm{v}_3^{(n)} + \mathrm{v}_5^{(n)})^6(3\mathrm{v}_2^{(n)} + \mathrm{v}_6^{(n)}) \\ &\sim \mathrm{v}_0^{(n)}(4\mathrm{v}_3^{(n)})^2(2\mathrm{v}_2^{(n)})^6(\mathrm{v}_1^{(n)})^6(3\mathrm{v}_2^{(n)}) = 3072(\mathrm{v}_0\mathrm{v}_1^6\mathrm{v}_2^7\mathrm{v}_3^2)^{(n)} \end{aligned}$$

Let F_2 be the fraction with defining relations $\emptyset = 123 = 3456 = 12456$. Looking at (12),

$$\begin{aligned} D(F_2) &= (\mathrm{v}_0^{(n)} + \mathrm{v}_3^{(n)} + \mathrm{v}_4^{(n)} + \mathrm{v}_5^{(n)})(\mathrm{v}_1^{(n)} + \mathrm{v}_2^{(n)} + \mathrm{v}_4^{(n)} + \mathrm{v}_5^{(n)})^2(\mathrm{v}_1^{(n)} + \mathrm{v}_3^{(n)} + 2\mathrm{v}_4^{(n)})^3 \\ &\qquad (\mathrm{v}_1^{(n)} + \mathrm{v}_2^{(n)} + \mathrm{v}_3^{(n)} + \mathrm{v}_6^{(n)})(\mathrm{v}_2^{(n)} + 2\mathrm{v}_3^{(n)} + \mathrm{v}_4^{(n)})^6(2\mathrm{v}_2^{(n)} + \mathrm{v}_3^{(n)} + \mathrm{v}_5^{(n)})^3 \\ &\sim \mathrm{v}_0^{(n)}(\mathrm{v}_1^{(n)})^2(\mathrm{v}_1^{(n)})(\mathrm{v}_1^{(n)})^3(\mathrm{v}_2^{(n)})^6(2\mathrm{v}_2^{(n)})^3 = 8(\mathrm{v}_0\mathrm{v}_1^6\mathrm{v}_2^9)^{(n)} \end{aligned}$$

Since $3072(\mathrm{v}_0\mathrm{v}_1^6\mathrm{v}_2^7\mathrm{v}_3^2)^{(n)}/8(\mathrm{v}_0\mathrm{v}_1^6\mathrm{v}_2^9)^{(n)} = 384(\mathrm{v}_3/\mathrm{v}_2)^{2n} \to 0$, F_2 is asymptotically better than F_1.

An intuitive explanation for the asymptotic superiority of F_2 over F_1 goes as follows: As $R^{(n)}$ approaches J, it becomes less and less important to learn about larger order effects relative to lower order effects. In order to gain the most independent information on small order interactions, they must be in different alias sets. With F_1, seven alias sets contain multiple two-factor interactions, while two alias sets have shortest word length 3. Design F_2 is asymptotically

better then F_1 because it sacrifices all three-factors effects in favor of providing more independent information about two-factor effects.

From a traditional, model-estimation point of view, this criterion favors designs that allow one to estimate (in the frequentist sense) models with the most low-order effects. It is similar to the *estimation capacity* criterion of Cheng, Steinberg and Sun (1999), which seeks to maximize the number of estimable models with all main effects and j two-way interactions.

7.2. Exponential near-dependence, observation with error

Next suppose the process with correlation $R^{(n)}$, as defined in Section 7.1, is observed with error. The asymptotically D-optimal fraction maximizes $\prod_{\mathcal{A}}(\sum_{W\in\mathcal{A}} \mathrm{v}_W^{(n)} + \sigma_p^2)$ as $n \to \infty$. Since $\mathrm{v}_\emptyset^{(n)} \to 1$ and $\mathrm{v}_W^{(n)} \to 0$ for all $W \neq \emptyset$, the principal terms after expanding this product are $\sigma_p^{2^{k-p+1}} + \sigma_p^{2(2^{k-p}-1)}\mathrm{v}_\emptyset^{(n)} + \sigma_p^{2(2^{k-p}-2)}\mathrm{v}_\emptyset^{(n)} \sum_{W\notin\mathcal{A}_F} \mathrm{v}_W^{(n)}$. So for large n the D-optimal design should maximize $\sum_{W\notin\mathcal{A}_F} \mathrm{v}_W^{(n)}$ or, equivalently, minimize $\sum_{W\in\mathcal{A}_F} \mathrm{v}_W^{(n)}$. As $n \to \infty$ the D-optimal design minimizes the maximum interaction variance among the defining words.

This criterion is more familiar in the isotropic case. Then the asymptotically D-optimal fraction minimizes $\sum_i L_i \mathrm{v}_i^{(n)}$. Since $\mathrm{v}_{i+1}^{(n)}/\mathrm{v}_i^{(n)} \to 0$ for orderly isotropic distributions and L_0 is always 1, the asymptotically D-optimal design maximizes the length of the shortest defining word in $\mathcal{A}_F$ and then minimizes the number of words of that length in $\mathcal{A}_F$. Thus the D-optimal design has maximum resolution and weak minimum aberration (Chen and Hedayat (1996)).

7.3. Linear near-dependence, observation without error

The next question is D-optimality for the correlation $R^\lambda = \lambda R + (1-\lambda)J$ as $\lambda \to 0$. It is again convenient to consider the cases of observation with and without error separately. Here we discuss error-free observation, deferring observation with error to Section 7.4.

The D-optimal fractional factorial maximizes $|\lambda R_F + (1-\lambda)J|$. The orthonormal matrix $2^{(k-p)/2}O_F$ simultaneously diagonalizes R_F and J: $R_F = (2^{(k-p)/2}O_F')2^{k-p}V_F(2^{(k-p)/2}O_F)$ and $J = (2^{(k-p)/2}O_F')2^{k-p}V_J(2^{(k-p)/2}O_F)$. As $\lambda \to 0$, interaction variances (except $\mathrm{v}_\emptyset$) approach 0 at the same rate, so we consider this a linear approach to dependence. A diagonal element of V_F is $\mathrm{v}_\mathcal{A} = \sum_{W\in\mathcal{A}} \mathrm{v}_W$, where $\mathcal{A}$ is an alias set for F. The matrix V_J is 0 everywhere except for a 1 in the diagonal position corresponding to $\mathrm{v}_{\mathcal{A}_F}$ Maximizing $|\lambda R_F + (1-\lambda)J|$ is then equivalent to maximizing $|\lambda V_F + (1-\lambda)V_J| = (\lambda \mathrm{v}_{\mathcal{A}_F} + 1 - \lambda)\prod_{\mathcal{A}\neq\mathcal{A}_F} \lambda \mathrm{v}_\mathcal{A} = \lambda^{|F|-1}\left[(1-\lambda)|V_F|/\mathrm{v}_{\mathcal{A}_F} + \lambda|V_F|\right]$. For λ close to 0 the D-optimal fractional factorial maximizes $|V_F|/\mathrm{v}_{\mathcal{A}_F} = \prod_{\mathcal{A}\neq\mathcal{A}_F} \mathrm{v}_\mathcal{A}$.

Remark. Unfortunately, maximizing $\prod_{\mathcal{A}\neq\mathcal{A}_F} \mathrm{v}_{\mathcal{A}}$ is no more manageable than the non-asymptotic criterion of maximizing $\prod_{\mathcal{A}} \mathrm{v}_{\mathcal{A}}$. However, changing the premise slightly yields a more pleasing result.

Suppose instead one starts with a correlation R for which the D-optimal fractional factorial F is known. Let F have alias sets $\mathcal{A}$ and let F' be any other fractional factorial with alias sets $\mathcal{A}'$. D-optimality of F means $|R_{F'}|/|R_F| = \prod'_{\mathcal{A}} \sum_{W\in\mathcal{A}'} \mathrm{v}_W / \prod_{\mathcal{A}} \sum_{W\in\mathcal{A}} \mathrm{v}_W \leq 1$. The question is the D-optimality of F for the correlation $R^\lambda = \lambda R + (1-\lambda)J$.

Diagonalizing all the matrices, the D-efficiency of F' relative to F for R^λ is

$$\frac{|R^\lambda_{F'}|}{|R^\lambda_F|} = \frac{(1-\lambda+\lambda\mathrm{v}_{\mathcal{A}_{F'}})\prod_{\mathcal{A}'\neq\mathcal{A}_{F'}} \lambda\mathrm{v}_{\mathcal{A}'}}{(1-\lambda+\lambda\mathrm{v}_{\mathcal{A}_F})\prod_{\mathcal{A}\neq\mathcal{A}_F} \lambda\mathrm{v}_{\mathcal{A}}} = \frac{(1-\lambda)\frac{|R_{F'}|}{\mathrm{v}_{\mathcal{A}_{F'}}} + \lambda|R_{F'}|}{(1-\lambda)\frac{|R_F|}{\mathrm{v}_{\mathcal{A}_F}} + \lambda|R_F|}.$$

Since $|R_{F'}|/|R_F| \leq 1$, a sufficient condition for $|R^\lambda_{F'}|/|R^\lambda_F| \leq 1$ is $\mathrm{v}_{\mathcal{A}_F} \leq \mathrm{v}_{\mathcal{A}_{F'}}$, i.e.,

$$\sum_{W\in\mathcal{A}_F} \mathrm{v}_W \leq \sum_{W\in\mathcal{A}_{F'}} \mathrm{v}_W. \tag{13}$$

In other words, F is D-optimal for the whole range of distributions R^λ, $\lambda \in (0,1]$, if (13) holds. Minimizing $\sum_{W\in\mathcal{A}_F} \mathrm{v}_W$ is the **c**-optimality criterion for minimizing the prediction error of $\sum_{\mathbf{t}} X(\mathbf{t})$. Further, $\sum_{W\in\mathcal{A}_F} \mathrm{v}_W = 2^{k-p}\sum_{\mathbf{f}\in F} r(\mathbf{f}) = 2^{k-p}\sum D_i r_i$ in the isotropic case. Notice this condition is very similar to (11).

7.4. Linear near-dependence, observation with error

Again, $R^\lambda = \lambda R + (1-\lambda)J$. We now seek to maximize $|\lambda R_F + (1-\lambda)J + \sigma^2 I|$ as $\lambda \to 0$. We apply the same tools as in Section 7.3. Since R_F, J, and I are simultaneously diagonalizable, it is equivalent to maximize the determinant of $\lambda V_F + (1-\lambda)V_J + \sigma_p^2 I$. So the asymptotically D-optimal design maximizes

$$\begin{aligned}
&(\lambda\mathrm{v}_{\mathcal{A}_F} + (1-\lambda) + \sigma_p^2) \prod_{\mathcal{A}\neq\mathcal{A}_F} (\lambda\mathrm{v}_{\mathcal{A}} + \sigma_p^2) \\
&= (\lambda(\mathrm{v}_{\mathcal{A}_F} - 1) + 1 + \sigma_p^2) \prod_{\mathcal{A}\neq\mathcal{A}_F} (\lambda\mathrm{v}_{\mathcal{A}} + \sigma_p^2) \\
&= (1+\sigma_p^2)\sigma_p^{2(|F|-1)} + \lambda\Big[(\mathrm{v}_{\mathcal{A}_F}-1)\sigma_p^{2(|F|-1)} + \sigma_p^{2(|F|-2)}(1+\sigma_p^2)\sum_{\mathcal{A}\neq\mathcal{A}_F} \mathrm{v}_{\mathcal{A}}\Big] + o(\lambda).
\end{aligned}$$

For sufficiently small λ, the D-optimal fractional factorial maximizes $\sigma_p^2(\mathrm{v}_{\mathcal{A}_F} - 1) + (1+\sigma_p^2)\sum_{\mathcal{A}\neq\mathcal{A}_F} \mathrm{v}_{\mathcal{A}}$. Ignoring the $-\sigma_p^2$ term, the design maximizes $(\sigma_p^2 + 1)\sum_{\mathcal{A}} \mathrm{v}_{\mathcal{A}} - \mathrm{v}_{\mathcal{A}_F}$. Since $\sum_{\mathcal{A}} \mathrm{v}_{\mathcal{A}} = \sum_W \mathrm{v}_W$ is constant over designs, the asymptotically optimal fractional factorial minimizes $\mathrm{v}_{\mathcal{A}_F} = \sum_{W\in\mathcal{A}_F} \mathrm{v}_W$. This is the

same criterion that arose in Section 4.4 and in the remark in Section 7.3. In the isotropic case, the asymptotically optimal design minimizes $\sum D_i r_i$.

8. Discussion of Asymptotic D-Optimality Criteria

Table 1 summarizes the asymptotic D-optimality criteria derived in Sections 6 and 7. For simplicity, the criteria are presented for orderly isotropic processes.

Table 1. D-Optimal Criteria for Orderly Isotropic Processes. Notation: $\hat{\mathrm{v}}_{\mathcal{A}}$ is the maximum variance of an interaction in alias set $\mathcal{A}$, $n_{\mathcal{A}}$ is the number of interactions in $\mathcal{A}$ with variance $\hat{\mathrm{v}}_{\mathcal{A}}$, D_i is the number of runs in the design distance i from $\mathbf{1}$, r_i is $r(\mathbf{f})$ when $|\mathbf{f}| = i$.

Asymptotic Distribution	**D-Optimality Criterion**
Independence: Direct Products	maximin distance
Independence: Convex Combinations	min $\sum D_i r_i^2$
Dependence: Matrix Powers	
$\sigma^2 = 0$	max $\prod_{\mathcal{A}} n_{\mathcal{A}} \hat{\mathrm{v}}_{\mathcal{A}}$
$\sigma^2 > 0$	maximin word length
Dependence: Convex Combinations	
$\sigma^2 = 0$	max $\prod_{\mathcal{A} \neq \mathcal{A}_F} \mathrm{v}_{\mathcal{A}}$
$\sigma^2 > 0$	min $\sum D_i r_i$

A nearly-independent process on T arising from direct product powers and a nearly-dependent process arising from matrix powers are analogous. For the nearly-independent process, correlations are brought exponentially toward 0, with the smallest correlations converging to 0 much faster than the larger correlations. For near dependence, interaction variances approach 0 at the same rate. It is not surprising, then, that the approach toward independence brings about a maximin criterion on runs in the design whereas the approach toward dependence brings about a maximin criterion on word lengths (when $\sigma^2 > 0$).

Studying tables of designs for various k and p has led to a conjecture about the relationship between the maximin distance criterion, the maximin word length criterion, and minimizing the design correlation, $\sum D_i r_i$. The conjecture pertains to tame processes, isotropic processes whose correlations $\mathbf{r}$ are completely monotone, as defined in Section 2.3. To illustrate the conjecture, refer to Table 2, which gives 2^{8-3} fractional factorials of resolution III or higher that use all 8 factors in the defining words according to Lemma 4.1. The table lists designs in decreasing order according to the maximin word length criterion, which is the same as the classical resolution–aberration criterion.

Take the first two designs in the table and compute $\sum D_i r_i$. Subtracting this quantity for the first design from corresponding quantity for the second design,

one has $(2r_2+8r_3+10r_4+8r_5+2r_6+r_8)-(r_2+10r_3+11r_4+4r_5+3r_6+2r_7) = D^4r_2+2D^4r_3+D^4r_4$. For a tame process, D^4r_2, D^4r_3, and D^4r_4 are positive, implying the first design always has smaller design correlation.

Table 2. 2^{8-3} Fractional Factorials. Notation: $\hat{\mathrm{v}}_{\mathcal{A}}$ is the maximum variance of an interaction in alias set $\mathcal{A}$, $n_{\mathcal{A}}$ is the number of interactions in $\mathcal{A}$ with variance $\hat{\mathrm{v}}_{\mathcal{A}}$.

Defining Words	Word Lengths / Design Distances / $\prod_{\mathcal{A}} n_{\mathcal{A}}\hat{\mathrm{v}}_{\mathcal{A}}$							
	1	2	3	4	5	6	7	8
1234	0	0	0	3	4	0	0	0
1256	0	1	10	11	4	3	2	0
13578	$2^6 3^3 4^1 \mathrm{v}_1^8 \mathrm{v}_2^{20} \mathrm{v}_3^3$							
1234	0	0	0	5	0	2	0	0
1567	0	2	8	10	8	2	0	1
123568	$2^9 3^2 4^4 5^4 \mathrm{v}_1^8 \mathrm{v}_2^{15} \mathrm{v}_3^8$							
1234	0	0	0	6	0	0	0	1
1256	0	4	0	22	0	4	0	1
12345678	$2^{12} 4^9 8^2 \mathrm{v}_1^8 \mathrm{v}_2^{13} \mathrm{v}_3^8 \mathrm{v}_4^2$							
123	0	0	1	2	3	1	0	0
1456	0	2	9	9	6	4	1	0
124578	$2^6 3^4 \mathrm{v}_1^8 \mathrm{v}_2^{19} \mathrm{v}_3^4$							
123	0	0	1	3	2	0	1	0
1456	0	3	6	11	8	1	2	0
1234578	$2^{10} 3^1 4^2 \mathrm{v}_1^8 \mathrm{v}_2^{17} \mathrm{v}_3^6$							
123	0	0	2	1	2	2	0	0
456	0	3	8	7	8	5	0	0
124578	$2^7 \mathrm{v}_1^8 \mathrm{v}_2^{19} \mathrm{v}_3^4$							
123	0	0	2	2	1	1	1	0
145	0	4	5	9	10	2	1	0
1245678	$2^5 3^6 \mathrm{v}_1^8 \mathrm{v}_2^{17} \mathrm{v}_3^6$							
123	0	0	2	2	2	0	0	1
145	0	5	0	19	0	7	0	0
12345678	$2^{11} \mathrm{v}_1^8 \mathrm{v}_2^{17} \mathrm{v}_3^6$							
123	0	0	3	1	0	2	1	0
456	0	5	4	7	12	3	0	0
1245678	$2^8 \mathrm{v}_1^8 \mathrm{v}_2^{17} \mathrm{v}_3^6$							

This sort of analysis holds for other designs. For example, comparing $\sum D_i r_i$ for the second and third designs in the list one has $(4r_2+22r_4+4r_6+r_8)-(2r_2+8r_3+10r_4+8r_5+2r_6+r_8) = 2D^4 r_2$, which is positive for completely monotone $\mathbf{r}$. So for a tame process the second design always has smaller design correlation than the third design (but not as small as the first design).

When does this fail? Compare the design correlation for the third and fourth designs in Table 2. We have $(4r_2+22r_4+4r_6+r_8)-(2r_2+9r_3+9r_4+6r_5+4r_6+r_7) = 2r_2-9r_3+13r_4-6r_5-r_7+r_8$. To show that this quantity is not consistently positive or negative for completely monotone $\mathbf{r}$, set $r_i = \rho^i$. For $\rho = 0.25$, this difference is positive (0.0293) but for $\rho = 0.75$ it is negative (-0.0158). Therefore, it cannot be written as the sum of difference operators on the r_i.

Similar observations for other values of k and p have led to the following conjecture.

Conjecture 1. Let F and F' be 2^{k-p} fractional factorials with respective design distances D_i and D_i' and word lengths L_i and L_i'. Then the following are equivalent:

(1) The design F is better than F' according to both the maximin word length criterion (resolution–aberration) and the maximin design distance criterion.

(2) The difference in design correlations, $\sum D_i r_i - \sum D_i' r_i = \sum (D_i - D_i') r_i$, can be written as a sum of difference operators on the r_i, $\sum_i \gamma_i D^s r_i$, with $\gamma_i \geq 0$. The difference operator D^s is such that s is the smallest i such that $L_i \neq L_i'$.

We conclude with examples applying asymptotic criteria to evaluate designs.

Example 3. Consider an experiment with eight binary factors for which an experimenter requests a recommendation for a one-eighth fraction. Assume the experimenter's prior knowledge is well-suited by a tame process on the design space.

Refer again to Table 2. Scanning the list, one sees the maximin word length design is also the maximin distance design. According to Conjecture 1 (and verified explicitly), the first design then also has the smallest design correlation $\sum D_i r_i$. Moreover, r_i^2 is completely monotone whenever r_i is (Kerr (1999), Lemma 4.4.1). Since the first design minimizes $\sum D_i r_i$ for all completely monotone sequences r_i, it also minimizes $\sum D_i r_i^2$.

In addition, one can check that this design is best according to maximizing $\prod_{\mathcal{A}} n_{\mathcal{A}} \hat{\mathrm{v}}_{\mathcal{A}}$ for any orderly isotropic prior distribution. For example, comparing the first and second designs, $2^9 3^2 4^4 5^4 (\mathrm{v}_1^8 \mathrm{v}_2^{15} \mathrm{v}_3^8)^n / 2^6 3^3 4^1 (\mathrm{v}_1^8 \mathrm{v}_2^{20} \mathrm{v}_3^3)^n = 2^9 5^4 \mathrm{v}_3^{5n} / 3 \mathrm{v}_2^{5n} \to 0$ as $n \to \infty$, so the first design is better. Altogether, for every criterion in Table 1 (ignoring $\prod_{\mathcal{A} \neq \mathcal{A}_F} \mathrm{v}_{\mathcal{A}}$, which we cannot evaluate), the first design is optimal. This gives one confidence in recommending this design to an experimenter because it is optimal for distributions at opposite extremes. That is, for both

nearly-dependent and nearly-independent distributions on T, the first design is optimal.

Example 4. Chen (1998) gives the maximum resolution–minimum aberration 2^{14-7} design (i.e., the maximin word length design), which is generated by $\emptyset = 1238 = 4569 = 1245\overline{10} = 1346\overline{11} = 1246\overline{12} = 2345\overline{14} = 1356\overline{13}$. The vector of design distances is the same as the vector of word lengths for this design. Therefore it is also the maximin distance design. In addition, if Conjecture 1 is correct then this design also minimizes $\sum D_i r_i$ and $\sum D_i r_i^2$ for tame processes among all 2^{14-7} fractional factorials.

Table 3. 2^{11-5} Fractional Factorials. Notation: $\hat{\mathrm{v}}_{\mathcal{A}}$ is the maximum variance of an interaction in alias set $\mathcal{A}$, $n_{\mathcal{A}}$ is the number of interactions in $\mathcal{A}$ with variance $\hat{\mathrm{v}}_{\mathcal{A}}$.

Defining Words	Word Lengths / Design Distances / $\prod_{\mathcal{A}} n_{\mathcal{A}} \hat{\mathrm{v}}_{\mathcal{A}}$										
	1	2	3	4	5	6	7	8	9	10	11
3457,2458	0	0	0	4	14	8	0	3	2	0	0
123469,123$\overline{10}$	0	0	2	14	22	8	6	9	2	0	0
1456$\overline{11}$	$2^{10}3^1 4^6 6^1 \mathrm{v}_1^{11} \mathrm{v}_2^{44} \mathrm{v}_3^8$										
34567,14568	0	0	0	5	10	10	5	0	0	0	1
12569, 1236$\overline{10}$	0	0	0	25	0	27	0	10	0	1	0
2346$\overline{11}$	$2^{15}5^{12} \mathrm{v}_1^{11} \mathrm{v}_2^{40} \mathrm{v}_3^{12}$										

Example 5. Table 3 gives two 2^{11-5} fractional factorials. The first design in Table 3 is the maximin word length fraction (Franklin (1980)), but the second design is better according to maximin distance. However, $[2^{15}5^{12}/(2^{10}3^1 4^6 6^1)]$ $[\mathrm{v}_1^{11}\mathrm{v}_2^{40}\mathrm{v}_3^{12}/(\mathrm{v}_1^{11}\mathrm{v}_2^{44}\mathrm{v}_3^8)]^n \sim [5^{12}/(2^8 3^2)][\mathrm{v}_3/\mathrm{v}_2]^{4n} \to 0$, so the first design beats the second design in maximizing $\prod_{\mathcal{A}} n_{\mathcal{A}} \hat{\mathrm{v}}_{\mathcal{A}}$.

Because the maximin word length design and the maximin distance designs are not the same, Conjecture 1 does not apply. Subtracting the design correlation $\sum D_i r_i$ for the second design from the first design, we have $2r_3 - 11r_4 + 22r_5 - 19r_6 + 6r_7 - r_8 + 2r_9 - r_{10}$. Figure 2 graphs this difference for the family of correlations $r_i = \rho^i$, with $0 \le \rho \le 1$. We see that for smaller values of ρ (approximately $\rho < 0.47$) the second design has smaller design correlation because this difference is positive. Smaller values of ρ mean more run-to-run independence for the process. For larger values of ρ (approximately $\rho > 0.47$), the first design has smaller design correlation and so is better based on this criterion.

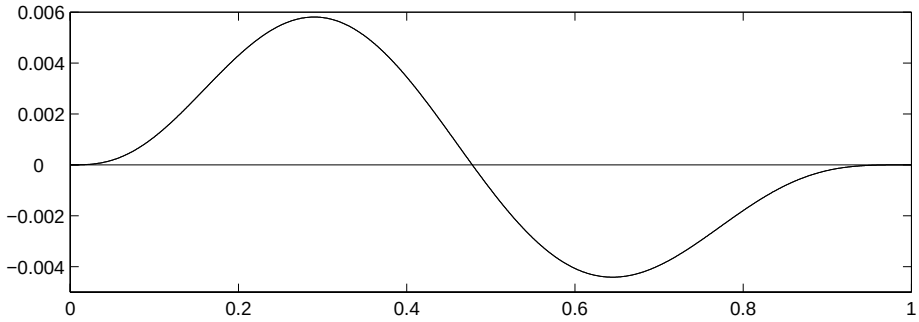

Figure 2. Difference of design correlations $\sum D_i\rho^i$ for two 2^{11-5} fractional factorials, as a function of ρ.

In this example, the first design is most appropriate for experiments modeled with a process with high run-to-run dependence because it has maximin word length, maximizes $\prod_{\mathcal{A}} n_{\mathcal{A}}\hat{\mathrm{v}}_{\mathcal{A}}$, and minimizes $\sum D_i\rho^i$ for ρ near 1. The second design is better for experiments modeled by nearly-independent processes, since it has maximin distance and minimizes $\sum D_i\rho^i$ for ρ near 0.

In our experience, the most common criterion for which the maximum resolution–minimum aberration design is sub-optimal is the criterion to maximize $\prod_{\mathcal{A}} n_{\mathcal{A}}\hat{\mathrm{v}}_{\mathcal{A}}$. The designs that are best for this criterion are the same designs proposed by other authors as improvements over maximum resolution–minimum aberration designs. For example, Chen, Sun and Wu (1993) note that the 2^{6-2} design called F_2 in Example 2 might be preferable to the maximum resolution–minimum aberration design because it allows independent estimates of some two-factor interactions. Altogether, the designs that appear in the Bayesian framework are reasonable.

Acknowledgement

I completed this research for my doctoral dissertation at UCLA under the mentorship of Don Ylvisaker, for whose guidance I am profoundly grateful. All errors and shortcomings are my own.

Appendix A. Proof of Lemma 4.1

We make a simple observation before discussing the proof.

Observation 1. *Let a, b, c, d be non-negative numbers and $a \geq b \geq c \geq d$. Then*

$$(a+d)(b+c) \geq (a+c)(b+d) \geq (a+b)(c+d).$$

Given four non-negative numbers $a \geq b \geq c \geq d$, in the context of Observation 1, refer to $(a+d)(b+c)$ as *best*, $(a+c)(b+d)$ as *middle*, and $(a+b)(c+d)$ as *worst*.

Let X be a stationary process on T with nested decreasing interaction variances and let F be a fractional factorial in which some factor is omitted from every defining word. Lemma 4.1 states that there exists a fraction of the same size F^+ such that $|R_{F^+}+\sigma^2 I| > |R_F+\sigma^2 I|$ for $\sigma^2 \geq 0$. Equivalently, $D(F^+) > D(F)$.

To begin the proof, first consider the case $\sigma^2 = 0$.

Let $W_1, \ldots, W_p$ be independent words generating $\mathcal{A}_F$ and L be a factor omitted from every word in $\mathcal{A}_F$. Choose any one of these words, say W_p, and consider the fractional factorial F^+ generated by $W_1, \ldots, W_{p-1}, W_pL$. Referring to the D-optimality criterion at (6), the proof will show $D(F^+) > D(F)$. In other words, we will compare $\prod_{\mathcal{A}} \mathrm{v}_{\mathcal{A}} = \prod_{\mathcal{A}} \sum_{W \in \mathcal{A}} \mathrm{v}_W$ for F and F^+.

The subgroup $< W_1, \ldots, W_p >= \mathcal{A}_F$ is all 2^p defining words for F. The defining words for F^+ are the subgroup $< W_1, \ldots, W_pL >= \mathcal{A}_{F^+}$. Construct a "super-group" S of 2^{p+1} words, $S =< W_1, \ldots, W_p, W_pL >$. Write all the words in S in the natural order according to the given listing of its generators:

$$S = \{ \overbrace{\emptyset, W_1, W_2, W_1W_2, \ldots, W_1W_2 \cdot\cdot W_{p-1}}^{S_1}, \overbrace{W_p, W_1W_p, \ldots, W_1 \cdot\cdot W_{p-1}W_p}^{S_2},$$
$$\underbrace{W_pL, W_1W_pL, \ldots, W_1 \cdot\cdot W_{p-1}W_pL}_{S_3}, \underbrace{W_pW_pL, W_1W_pW_pL, \ldots, W_1 \cdot\cdot W_pW_pL}_{S_4} \}$$

Notice the first half of S is $\mathcal{A}_F$ and the second half is one alias set for F. Similarly, the first quarter and the third quarter are the defining relations $\mathcal{A}_{F^+}$ and the second and fourth quarters together are alias set for F^+. Calling the *ith* quarter S_i, one can write $S = S_1 \cup S_2 \cup S_3 \cup S_4$, $\mathcal{A}_F = S_1 \cup S_2$, $\mathcal{A}_{F^+} = S_1 \cup S_3$, $W_pL\mathcal{A}_F = L\mathcal{A}_F = S_3 \cup S_4$, $W_p\mathcal{A}_{F^+} = L\mathcal{A}_{F^+} = S_2 \cup S_4$.

The strategy is to use S and its cosets in the word group to match pairs of factors in the product for $D(F)$ with pairs of factors in the product for $D(F^+)$. Write $\mathrm{v}_{S_i} = \sum_{W \in S_i} \mathrm{v}_W$ and v_{US_i} analogously for translates of S_i.

Two factors in the product for $D(F)$ are $(\mathrm{v}_{S_1} + \mathrm{v}_{S_2})(\mathrm{v}_{S_3} + \mathrm{v}_{S_4})$ and two factors in the product for $D(F^+)$ are $(\mathrm{v}_{S_1} + \mathrm{v}_{S_3})(\mathrm{v}_{S_2} + \mathrm{v}_{S_4})$. Notice that for every word $W \in S_1$ there is the word $WL \in S_4$. When $L \notin W$, $W \subset WL$, so $\mathrm{v}_W > \mathrm{v}_{WL}$. Similarly, every word $W \in S_2$ does not contain L so $W \subset WL \in S_3 \Longrightarrow \mathrm{v}_W > \mathrm{v}_{WL}$. Therefore $\mathrm{v}_{S_1} > \mathrm{v}_{S_4}$ and $\mathrm{v}_{S_2} > \mathrm{v}_{S_3}$. Using the terminology following Observation 1, the arrangement of $\mathrm{v}_1, \mathrm{v}_2, \mathrm{v}_3, \mathrm{v}_4$ given by F cannot be best. If it is worst then the arrangement given by F^+ is middle or best – in either case, better than F. If the arrangement of $\mathrm{v}_1, \mathrm{v}_2, \mathrm{v}_3, \mathrm{v}_4$ given by F is middle, then either (i) $\mathrm{v}_{S_1} > \mathrm{v}_{S_4} > \mathrm{v}_{S_2} > \mathrm{v}_{S_3}$ or (ii) $\mathrm{v}_{S_2} > \mathrm{v}_{S_3} > \mathrm{v}_{S_1} > \mathrm{v}_{S_4}$. But both

(i) and (ii) imply the arrangement given by F^+ is best. Altogether, conclude $(\mathrm{v}_{S_1} + \mathrm{v}_{S_3})(\mathrm{v}_{S_2} + \mathrm{v}_{S_4}) > (\mathrm{v}_{S_1} + \mathrm{v}_{S_2})(\mathrm{v}_{S_3} + \mathrm{v}_{S_4})$

Now consider any coset of S, $US = US_1 \cup US_2 \cup US_3 \cup US_4$, for any word $U \notin S$. Once again, the product for $D(F)$ contains $(\mathrm{v}_{US_1} + \mathrm{v}_{US_2})(\mathrm{v}_{US_3} + \mathrm{v}_{US_4})$, whereas the product for $D(F^+)$ contains $(\mathrm{v}_{US_1} + \mathrm{v}_{US_3})(\mathrm{v}_{US_2} + \mathrm{v}_{US_4})$. There are two cases, $L \notin U$ and $L \in U$.

If $L \notin U$, repeat the argument above: for every word $UW \in US_1$ there is the word $UWL \in US_4$ and $UW \subset UWL \Longrightarrow \mathrm{v}_{UW} > \mathrm{v}_{UWL}$, and so on.

If $L \in U$ only a minor adjustment to the argument is needed. For every word $UW \in US_1$, $UWL \in US_4$, $UW \supset UWL$. For every word $UW \in US_2$ $UWL \in US_3$, $UW \supset UWL$. So $\mathrm{v}_{US_1} < \mathrm{v}_{US_4}$ and $\mathrm{v}_{US_2} < \mathrm{v}_{US_3}$. This is enough to know the arrangement of $\mathrm{v}_{US_1}, \mathrm{v}_{US_2}, \mathrm{v}_{US_3}, \mathrm{v}_{US_4}$ given by F is not best. If it is worst, then the arrangement given by F^+ is better (either middle or best). If the arrangement of $\mathrm{v}_{US_1}, \mathrm{v}_{US_2}, \mathrm{v}_{US_3}, \mathrm{v}_{US_4}$ given by F is middle, then either (i) $\mathrm{v}_{US_1} < \mathrm{v}_{US_4} < \mathrm{v}_{US_2} < \mathrm{v}_{US_3}$ or (ii) $\mathrm{v}_{US_2} < \mathrm{v}_{US_3} < \mathrm{v}_{US_1} < \mathrm{v}_{US_4}$. Both (i) and (ii) imply the arrangement given by F^+ is then best.

In summary, S matches every pair of factors in $\prod_{\mathcal{A}} \mathrm{v}_{\mathcal{A}}$ uniquely to a pair of factors in $\prod_{\mathcal{A}^+} \mathrm{v}_{\mathcal{A}^+}$ that have a larger product. Therefore $D(F) < D(F^+)$. In particular, F is not D-optimal.

Finally, if $\sigma^2 > 0$ write $\mathrm{v}'_W = \mathrm{v}_W + \sigma_p^2/2^p$. Since the v_W are nested decreasing, the v'_W are nested decreasing. Since D-optimality means maximizing $\prod_{\mathcal{A}}(\sigma_p^2 + \sum_{W \in \mathcal{A}} \mathrm{v}_W) = \prod_{\mathcal{A}} \sum_{W \in \mathcal{A}} \mathrm{v}'_W$, the same proof holds using v'_W in place of v_W.

Appendix B. Proof of Lemma 5.1

The proof of Lemma 5.1 requires a second observation.

Observation 2. *Let a, b, c, d, and s be non-negative numbers and $a \geq b \geq c \geq d$. Then*

$$\frac{a^2+d^2}{a+d+s} + \frac{b^2+c^2}{b+c+s} \geq \frac{a^2+c^2}{a+c+s} + \frac{b^2+d^2}{b+d+s} \geq \frac{a^2+b^2}{a+b+s} + \frac{c^2+d^2}{c+d+s}.$$

In the context of Observation 2, refer to $(a^2+d^2)/(a+d+s)+(b^2+c^2)/(b+c+s)$ as *best*, $(a^2+c^2)/(a+c+s)+(b^2+d^2)/(b+d+s)$ as *middle*, $(a^2+b^2)/(a+b+s)+(c^2+d^2)/(c+d+s)$ as *worst*.

Let X be a stationary process on T with nested decreasing interaction variances. Let W_l be any word of length l, $2 \leq l \leq k$, W_{l-1} be a word of length $l-1$, $W_{l-1} \subset W_l$. Let F be the half-fraction defined by $\emptyset = W_l$ and F' be the half-fraction defined by $\emptyset = W_{l-1}$. Lemma 5.1 states that $\sum_{\mathbf{t}} \mathrm{Var}_F(X(\mathbf{t})) < \sum_{\mathbf{t}} \mathrm{Var}_{F'}(X(\mathbf{t}))$, i.e., the total posterior variance is smaller for X conditioned on F than on F'.

Write $W_l = LW_{l-1}$, i.e., W_{l-1} is just W_L without factor L. Referring to the A-optimality criterion at (7), the proof shows $A(F) > A(F')$.

Partition the group of words into cosets of the subgroup $S = \{\emptyset, W_{l-1}, W_l, L\}$. Notice that if a word U is contained in W_l then all the words in the coset US are contained in W_l.

First consider cosets where all four words are contained in W_l. In such a coset, exactly two of the words are contained in W_{l-1}: (i) if $U \subseteq W_{l-1}$, then $W_{l-1}U \subseteq W_{l-1}$ but $L \in W_lU$ and $L \in W_lW_{l-1}U$; (ii) if $U \not\subseteq W_{l-1}$, then $L \in U$, $L \in W_{l-1}U$ and $L \notin W_lU, W_lW_{l-1}U$. Without loss of generality, assume $U \subseteq W_{l-1}$.

Now, $A(F)$ contains the terms $(\mathrm{v}_U^2 + \mathrm{v}_{UW_l}^2)/(\mathrm{v}_U + \mathrm{v}_{UW_l} + \sigma_p^2) + (\mathrm{v}_{UW_{l-1}}^2 + \mathrm{v}_{UL}^2)/(\mathrm{v}_{UW_{l-1}} + \mathrm{v}_{UL} + \sigma_p^2)$. The corresponding terms for $A(F')$ are $(\mathrm{v}_U^2 + \mathrm{v}_{UW_{l-1}}^2)/(\mathrm{v}_U + \mathrm{v}_{UW_{l-1}} + \sigma_p^2) + (\mathrm{v}_{UW_l}^2 + \mathrm{v}_{UL}^2)/(\mathrm{v}_{UW_l} + \mathrm{v}_{UL} + \sigma_p^2)$.

Since $U \subset LU$ and $W_{l-1}U \subset LW_{l-1}U$, $\mathrm{v}_U > \mathrm{v}_{LU}$ and $\mathrm{v}_{W_{l-1}U} > \mathrm{v}_{W_lU}$. This is enough to know the arrangement of $\mathrm{v}_U, \mathrm{v}_{W_{l-1}U}, \mathrm{v}_{W_lU}, \mathrm{v}_{LU}$ given by F' is not best, so it is either middle or worst. If it is worst, the arrangement given by F is better – either middle or best. If the arrangement given by F' is middle, then either (i) $\mathrm{v}_U > \mathrm{v}_{LU} > \mathrm{v}_{W_{l-1}U} > \mathrm{v}_{W_lU}$, or (ii) $\mathrm{v}_{W_{l-1}U} > \mathrm{v}_{W_lU} > \mathrm{v}_U > \mathrm{v}_{LU}$. Both (i) and (ii) imply that the arrangement of $\mathrm{v}_U, \mathrm{v}_{W_{l-1}U}, \mathrm{v}_{W_lU}, \mathrm{v}_{LU}$ given by F is best.

Finally, consider the remaining cosets of S where none of the coset members are contained in W_l. For U in such a coset, one can write $U = U_0U_1$, where $U_0 \subseteq W_l$ (possibly the empty set) and U_1 nonempty, $U_1 \cap W_l = \emptyset$. Then consider the four words U, W_lU, $W_{l-1}U$, $W_lW_{l-1}U = LU$, which can be written U_0U_1, $W_lU_0U_1$, $W_{l-1}U_0U_1$, $W_lW_{l-1}U_0U_1 = LU_0U_1$. Since U_1 is disjoint from W_l and so also from U_0 and W_{l-1}, adjoining U_1 to the four words U_0, W_lU_0, $W_{l-1}U_0$, $W_lW_{l-1}U_0 = LU_0$ does not change any subset or superset relations. So ignore U_1 and consider this case in terms of U_0. Since $U_0 \subseteq W_l$, this reduces to the previous case.

References

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

Chen, H. and Hedayat, A. S. (1996). 2^{n-l} Designs with weak minimum aberration. *Ann. Statist.* **24**, 2526-2548.

Chen, J. (1998). Intelligent search for 2^{13-6} and 2^{14-7} minimum aberration designs. *Statist. Sinica* **8**, 1265-1270.

Chen, J., Sun, D. X. and Wu, C. F. J. (1993). A catalogue of two-level and three-level fractional factorial designs with small runs. *Internat. Statist. Rev.* **61**, 131-145.

Chen, J. and Wu, C. F. J. (1991). Some results on s^{n-k} fractional factorial designs with minimum aberration or optimal moments. *Ann. Statist.* **19**, 1028-1041.

DuMouchel, W. and Jones, B. (1994). A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics* **36**, 37-47.

Franklin, M. F. (1985). Selecting defining contrasts and confounded effects in p^{n-m} factorial experiments. *Technometrics* **27**, 165-172.

Fries, A. and Hunter, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics* **22**, 601-608.

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

Kerr, M. K. (1999). Stationary process on 2^k for Bayesian experimental design. Ph. D. Thesis, University of California, Los Angeles.

Laycock, P. J. and Rowley P. J. (1995). A method for generating and labelling all regular fractions or blocks for q^{n-m}-designs. *J. Roy. Statist. Soc.* **57**, 191-204.

Meyer, R. D., Steinberg, D. M. and Box, G. (1996). Follow-up designs to resolve confounding in multifactor experiments. *Technometrics* **38**, 303-313.

Mitchell, T. J., Morris, M. D. and Ylvisaker, D. (1995). Two-level fractional factorials and Bayesian prediction. *Statist. Sinica* **5**, 559-573.

Mitchell, T., Sacks, J. and Ylvisaker, D. (1994). Asymptotic Bayes criteria for nonparametric response surface design. *Ann. Statist.* **22**, 634-651.

Tang, B. and Wu, C. F. J. (1996). Characterization of minimum aberration 2^{n-k} designs in terms of their complementary designs. *Ann. Statist.* **24**, 2549-2559.

Toman, B. (1994). Bayes optimal designs for two- and three-level factorial experiments. *J. Amer. Statist. Assoc.* **89**, 937-946.

The Jackson Laboratory, 600 Main Street, Box 303, Bar Harbor, ME 04609, U.S.A.

E-mail: mkk@jax.org

(Received July 1999; accepted November 2000)