University of California

Los Angeles

Stationary Processes on 2^k for Bayesian Experimental Design

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Statistics

by

M. Kathleen Kerr

1999

The dissertation of M. Kathleen Kerr is approved.

Jan De Leeuw

Ker-Chau Li

Weng-Kee Wong

Don Ylvisaker, Committee Chair

University of California, Los Angeles

1999

For my family.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I thank my advisor, Don Ylvisaker, for his support and encouragement over the past several years, his guidance on this thesis, and especially for helping me know the kind of statistician I aspire to become.

Vita

1970	Born, Columbus, Ohio
1993	B.A. Magna Cum Laude, With Honors, Mathematics Bryn Mawr College
1993	Scott Prize in Mathematics Bryn Mawr College
1993-1996	Harris Fellowship University of California, Los Angeles
1994	M.A., Mathematics University of California, Los Angeles
1996	Hoel Scholar, Division of Statistics University of California, Los Angeles
1997-1999	AIDS Training Fellowship, Department of Biostatistics University of California, Los Angeles
1998	M.S., Statistics University of California, Los Angeles

Publications

Charles R. Johnson, M. Kathleen Kerr, David P. Stanford, *Semipositivity of Matrices.* Journal of Linear and Multilinear Algebra, 1994, **37**, 265-271.

Abstract of the Dissertation

Stationary Processes on 2^k for Bayesian Experimental Design

by

M. Kathleen Kerr

Doctor of Philosophy in Statistics

University of California, Los Angeles, 1999

Professor Don Ylvisaker, Chair

We consider an experiment with k two-level factors and insufficient resources to observe all 2^k possible runs. The candidate designs are 2^{k-p} fractional factorials. To approach the problem, we do not assume a parametric model and instead think of experimental observations as realizations of a stationary Gaussian process X operating on the design space. Pre-experimental knowledge is formally incorporated in the prior distribution of X, making the approach Bayesian. However, instead of demanding a precise prior for X, we seek designs that are optimal for general families of processes. We define several families of interest and prove each family is closed under various operations.

In evaluating designs, we examine criteria such as D-, A-, G-, E-, and c-optimality, paying closest attention to D-optimality. Within a family of processes, we consider different ways to bring a distribution towards near-independence and near-dependence, then characterize the asymptotically optimal fractional factorial. 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 priors.

CHAPTER 1

Introduction

We consider stationary Gaussian processes operating on the vertices of the k-dimensional cube T. Stationary processes have an interesting invariance with respect to the geometry and algebraic structure of T. Specifically, the covariance structure of a stationary process behaves nicely with respect to the group structure of T. Because of its special covariance structure, any stationary Gaussian process on T can be transformed and represented in terms of independent "interactions."

Following the lead of Mitchell, Morris, and Ylvisaker (1995), these processes provide a framework for thinking about experimental design in the situation of k two-level factors. The design space in this setting is T. One imagines experimental observations as realizations of a Gaussian process X operating on the design space. Pre-experimental knowledge is incorporated as a prior distribution for X. The process X is assumed to be stationary to ensure a kind of "impartiality" for the distribution of X.

This strategy for design is Bayesian in its use of prior knowledge. However, this approach differs from most previous Bayesian design work, which largely discusses estimating 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.

Besides Mitchell, Morris, and Ylvisaker (1995), the most closely related work is that of Toman (1994), who also considers designs for multiple two- and three-level factors by imagining observations as realizations of a stochastic process. In that paper the covariance matrix for the process on the design space is $\sigma^{-2}R$, where $R = (\omega^2 - \delta)I + \delta J$. This is the assumption of exchangeability and is a special case of the processes we consider. The proposed design criteria relate to minimizing the posterior variances of linear functions of responses. Classical 2^{k-p} 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).

This paper includes the case of error-free experimentation. Here, computer experiments are the canonical example. Although this setting is deterministic, design questions arise when computing is too time-consuming or expensive to explore the entire range of inputs. The stochastic process is meant to describe one's limited knowledge of the deterministic function on the design space (Currin, Mitchell, Morris, and Ylvisaker (1991)). We share the view in Sacks, Welch, Mitchell, and Wynn (1989) that a statistical framework is appropriate and useful for quantifying uncertainty in complex deterministic settings.

Chapter 2 establishes notation, recalls previous results, and generally sets the

groundwork. Chapter 3 begins to study stationary processes on T by defining and discussing three families of stationary processes of special interest. These families are studied further in Chapter 4, where they are shown to have pleasant closure properties. In Chapter 5 we generalize results for stationary processes on T to stationary processes restricted to subgroups of T, and derive distributions for processes conditioned on subgroups.

Since subgroups of T are the classical fractional factorial designs, these discussions lead naturally to discussing design. We define Bayesian criteria for good designs in Chapter 6, and this is followed by some basic results about optimal designs in Chapter 7. Drawing on the results in Chapter 4 about closure properties, Chapter 8 explores certain families of processes and characterizes some asymptotically optimal designs.

CHAPTER 2

Setup

2.1 The Space T

The space under study is $T = \{-1, 1\}^k$, all 2^k vectors with entries "1" and "-1." In other words, T is the vertices of the k-dimensional cube.

The space T has various mathematical properties. First, T is a metric space under Hamming distance d, where $d(\mathbf{t}, \mathbf{s})$ is the number of components in which $\mathbf{t}$ and $\mathbf{s}$ disagree. Second, T forms a group via component-wise direct multiplication with identity $\mathbf{1}$ and every element self-inverse. Hamming distance is invariant under translation by any group element $\mathbf{u} \in T$, i.e. $d(\mathbf{t}, \mathbf{s}) = d(\mathbf{tu}, \mathbf{su})$. Third, 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. The usual inner product, with arithmetic in $\mathbb{F}_2$, provides a notion of orthogonality.

Looking ahead, this space arises in experimental design, which motivates much of this paper. Consider an experiment with k variables, each of which has two levels. Denoting the two levels with "1" and "-1," the 2^k possible experimental combinations are the elements of T. Borrowing design terminology, the elements of T are called *runs* and the indices of the k components of a run are called *factors*. One thinks of functions on T as experimental responses for the different factor combinations.

2.2 Words, Functions on T, and Interactions

Use some set of symbols to represent the k factors defining T, say $\{1, 2, \ldots, k\}$. Subsets of $\{1, 2, \ldots, k\}$ are called *words* and the collection of words is denoted $\mathcal{W}$. A word $W \in \mathcal{W}$ has *length* $|W|$.

The set of words $\mathcal{W}$ forms 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 in the word group 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 is the same as binary addition on these vectors. This notation makes it obvious that $\mathcal{W}$ and T are isomorphic groups. In particular, like T, $\mathcal{W}$ forms a vector space over $\mathbb{F}_2$, so it makes sense to discuss the (linear) independence of a set of words.

Every word W gives a function $T \to \{-1, 1\}$ defined by $W(\mathbf{t}) = \prod_{i \in W} t_i$. Because $W(\mathbf{s})W(\mathbf{t}) = W(\mathbf{st})$, these functions are actually homomorphisms on the group T. In addition, one can check that $W(\mathbf{t})U(\mathbf{t}) = WU(\mathbf{t})$. Thus, when thought of as functions, words preserve their group structure.

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. In the context of experimental design, interactions are important because they correspond to factor "effects." More generally, interactions have a useful mathematical role in studying processes on T.

2.3 Stationary Processes on T

Consider a mean zero, Gaussian processes on T. In our notation, X is such a process that is available for observation (although possibly only 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 $Z(\mathbf{t}) = X(\mathbf{ut})$. The stationarity assumption imposes the constraint that the relationship between observations $X(\mathbf{t})$ and $X(\mathbf{s})$ depends only on the vector $\mathbf{ts}$ that indicates the factors in which $\mathbf{t}$ and $\mathbf{s}$ differ. For example, when $\mathbf{t}$ and $\mathbf{s}$ are adjacent vertices on the k-cube, the correlation between $X(\mathbf{t})$ and $X(\mathbf{s})$ is the same as the correlation between $X(\mathbf{tu})$ and $X(\mathbf{su})$. The runs $\mathbf{tu}$ and $\mathbf{su}$ are again an adjacent pair of vertices, and their edge is parallel to the edge between $\mathbf{t}$ and $\mathbf{s}$.

A more general setup is

$$X(\mathbf{t}) = \sum_{W} W(\mathbf{t})\beta_W + Y(\mathbf{t}). \tag{2.1}$$

The β_W are independent, mean-zero normal random variables with variance ν_W. Normally, we think of $\nu_W = 0$ for $|W|$ large so that we are extracting "global" effects for low-order interactions. In (2.1) Y is a mean-zero stationary Gaussian process and is meant to represent the "local" process.

In (2.1), we assume Y is stationary, which automatically implies stationarity for X. To see this, first note that that the sum of stationary processes is again a stationary process. So let $\mathbf{u}$ be any fixed run in T and consider the joint

distribution of $\{\sum_W W(\mathbf{t})\beta_W\}$ for $\mathbf{t} \in T$. Then

$$\begin{aligned} E\left[\sum_W W(\mathbf{s})\beta_W \sum_W W(\mathbf{t})\beta_W\right] &= \sum_W W(\mathbf{s})W(\mathbf{t})\nu_W \\ &= \sum_W W(\mathbf{s})W(\mathbf{t})W^2(\mathbf{u})\nu_W \\ &= \sum_W W(\mathbf{su})W(\mathbf{tu})\nu_W \\ &= E\left[\sum_W W(\mathbf{su})\beta_W \sum_W W(\mathbf{tu})\beta_W\right]. \end{aligned}$$

Therefore $\sum_W W(\mathbf{t})\beta_W$ is also a mean-zero stationary Gaussian process, implying X is as well.

When there is no observational error, observations on T are thought of as realizations of the process X. In the presence of observational error, the i^{th} observation at $\mathbf{t} \in T$ is $X_i(\mathbf{t})$, where

$$X_i(\mathbf{t}) = X(\mathbf{t}) + \epsilon_i(\mathbf{t}). \tag{2.2}$$

Again, $\epsilon_i(\mathbf{t})$ is assumed to be a mean-zero Gaussian process. Further assume the error process is independent of X, and $E[\epsilon_i(\mathbf{t})\epsilon_j(\mathbf{s})] = \sigma^2\delta_{ij}\delta_{\mathbf{ts}}$. Finally, notice that independence among $\epsilon(\mathbf{t})$ is a special case of stationarity, so for fixed i, X_i in (2.2) is stationary.

2.3.1 Interactions for Stationary Processes

A stationary process X on T is a random function on T, so X has interactions as defined in Section 2.2. 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. Stationarity

simplifies the expression for the interaction variances as follows:

$$\begin{aligned}
\mathrm{var}(\Gamma_W(X)) &= 2^{-2k} E[\sum_{\mathbf{s}} W(\mathbf{s})X(\mathbf{s}) \sum_{\mathbf{t}} W(\mathbf{t})X(\mathbf{t})] \\
&= 2^{-2k} \sum_{\mathbf{s}} \sum_{\mathbf{t}} W(\mathbf{s})W(\mathbf{t})E[X(\mathbf{s})X(\mathbf{t})] \\
&= 2^{-2k} \sum_{\mathbf{s}} \sum_{\mathbf{t}} W(\mathbf{st})E[X(\mathbf{1})X(\mathbf{st})] \\
&= 2^{-2k} \sum_{\mathbf{s}} \sum_{\mathbf{t}} W(\mathbf{t})E[X(\mathbf{1})X(\mathbf{t})] \\
&= 2^{-k} \sum_{\mathbf{t}} W(\mathbf{t})E[X(\mathbf{1})X(\mathbf{t})]. \qquad (2.3)
\end{aligned}$$

Similarly, the covariances of the process can be recovered from the interaction variances. Since $E[X(\mathbf{s})X(\mathbf{t})] = E[X(\mathbf{1})X(\mathbf{st})]$, for simplicity use $\mathbf{1}$ as one of the runs:

$$\begin{aligned}
E[X(\mathbf{1})X(\mathbf{t})] &= E[\sum_{U} \Gamma_U(X) \sum_{W} W(\mathbf{t})\Gamma_W(X)] \\
&= \sum_{W} W(\mathbf{t}) \sum_{U} E[\Gamma_U(X)\Gamma_W(X)] \\
&= \sum_{W} W(\mathbf{t})\mathrm{var}(\Gamma_W(X)). \qquad (2.4)
\end{aligned}$$

To ease notation, let $\mathrm{v}_W = \mathrm{var}(\Gamma_W(X))$. In the more general setup (2.1), let $\mathrm{v}_W^Y = \mathrm{var}(\Gamma_W(Y))$ and notice $\mathrm{v}_W = \nu_W + \mathrm{v}_W^Y$. For fixed i it makes mathematical sense to talk about $\Gamma_W(X_i)$ in (2.2), and $\mathrm{var}(\Gamma_W(X_i)) = \mathrm{v}_w + 2^{-k}\sigma^2$.

Example 1 (Mitchell, Morris, and Ylvisaker (1995)). Start with k positive correlations $\rho_1, \rho_2, \ldots, \rho_k$ and set $E[X(\mathbf{s})X(\mathbf{t})] = \prod_{i:s_i \neq t_i} \rho_i$. According to (2.3), $\mathrm{v}_W = 2^{-k} \sum_{\mathbf{t}} W(\mathbf{t})E[X(\mathbf{1})X(\mathbf{t})] = 2^{-k} \sum_{\mathbf{t}} W(\mathbf{t}) \prod_{i:t_i=-1} \rho_i = 2^{-k} \prod_{i \in W}(1 - \rho_i) \prod_{i \notin W}(1 + \rho_i)$.

A special case has $\rho_1 = \rho_2 = \ldots = \rho_k = \rho$. Then $E[X(\mathbf{s})X(\mathbf{t})] = \prod_{i:s_i \neq t_i} \rho = \rho^{d(\mathbf{s},\mathbf{t})}$ and $\mathrm{v}_W = 2^{-k}(1-\rho)^{|W|}(1+\rho)^{k-|W|}$. □

2.3.2 Covariance Matrices for Stationary Processes

The 2^k random variables $X(\mathbf{t})$ for the stationary process X can be transformed to 2^k *independent* interactions Γ_W. We summarize this transformation using matrix notation. Write Γ for the vector of 2^k interactions, $\mathbf{X}$ for the 2^k vector of $X(\mathbf{t})$, and $O = O_k$ for the $2^k \times 2^k$ matrix that transforms $\mathbf{X}$ to Γ. Then $\Gamma = O\mathbf{X}$. Next consider the covariance matrix R for the process X on T. Since $\Gamma = O\mathbf{X}$, $V = \mathrm{var}(\Gamma) = ORO'$. Because the interactions for a stationary process are independent, V is a diagonal matrix with the interaction variances v_W as its diagonal entries. Moreover, one can check from the definition of the interactions as contrasts of the $\{X(\mathbf{t})\}$ that O is an orthogonal matrix. To see this, notice (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 the diagonalization of R. Since R is symmetric, 2^kV gives the eigenvalues of R. In other words, the eigenvalues of R are nothing more than the interaction variances for the process, multiplied by 2^k.

Example 2 Set $k = 2$ so that $T = \{(1,1),(-1,1),(1,-1),(-1,-1)\} = \{\mathbf{1}, \mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_{12}\}$. Let X be a stationary process on T. In complete generality, let $\mathrm{var}(X(\mathbf{t})) = 1$ and R_* denote $E[X(\mathbf{1})X(\mathbf{t}_*)]$. Then

$$R = \begin{bmatrix} 1 & R_1 & R_2 & R_{12} \\ R_1 & 1 & R_{12} & R_2 \\ R_2 & R_{12} & 1 & R_1 \\ R_{12} & R_2 & R_1 & 1 \end{bmatrix},$$

and, using (2.3), the interaction variances are

$$\begin{aligned}
\mathrm{v}_\emptyset &= \frac{1}{4}(1 + R_1 + R_2 + R_{12}), \\
\mathrm{v}_1 &= \frac{1}{4}(1 - R_1 + R_2 - R_{12}), \\
\mathrm{v}_2 &= \frac{1}{4}(1 + R_1 - R_2 - R_{12}), \\
\mathrm{v}_{12} &= \frac{1}{4}(1 - R_1 - R_2 + R_{12}).
\end{aligned}$$

In other words, for the orthogonal matrix

$$O = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

ORO' is diagonal with $\mathrm{v}_\emptyset, \mathrm{v}_1, \mathrm{v}_2$, and v_{12} along the diagonal. Since the v_W are proportional to the eigenvalues of R_*, R is positive definite as long as $1 + R_1 + R_2 + R_{12}, 1 - R_1 + R_2 - R_{12}, 1 + R_1 - R_2 - R_{12}$, and $1 - R_1 - R_2 + R_{12}$ are all positive. □

2.3.3 Dual Processes

Because of stationarity, the distribution of X on T is completely specified by 2^k covariances $R(\mathbf{1}, \mathbf{t})$ or by 2^k interaction variances v_W. Any process X such that $R(\mathbf{1}, \mathbf{t}) \geq 0$ for all $\mathbf{t}$ has a *dual process* X^*. The interaction variances for X become the covariances for X^* and the interaction variances for X^* are proportional to the covariances for X.

Let $T^* = \{-1, 1\}^k$ be another "copy" of T, also indexed by the k factors. We define a stationary process X^* on T^*. Similarly, let $\mathcal{W}^*$ be another "copy" of $\mathcal{W}$. As noted in Section 2.2, T, $\mathcal{W}$, T^*, and $\mathcal{W}^*$ are all isomorphic. Write the elements $\mathbf{t}^*$ of T^* as $\mathbf{t}^*_W$ to record their corresponding word in $\mathcal{W}$.

Begin constructing a covariance function R^* for the process X^* on T^* by defining $R^*(\mathbf{1}^*, \mathbf{t}^*_W) = \mathrm{v}_W$. For any $\mathbf{s}^*$ and $\mathbf{t}^*$ define $R^*(\mathbf{s}^*, \mathbf{t}^*) = R^*(\mathbf{1}^*, \mathbf{s}^*\mathbf{t}^*)$, and notice this definition guarantees stationarity for X^*.

Write the words for the dual process as W^* or $W^*_{\mathbf{t}}$, where W^* and $\mathbf{t}$ correspond according to the isomorphism between $\mathcal{W}^*$ and T. From our definition of the covariance structure for X^*, we can derive the variance for an interaction $\Gamma_{W^*_{\mathbf{t}}}(X^*)$:

$$\begin{aligned}
\mathrm{v}_{W^*_{\mathbf{t}}} &= 2^{-2k} E[\sum_{\mathbf{t}^*_W} W^*_{\mathbf{t}}(\mathbf{t}^*_W) X^*(\mathbf{t}^*_W) \sum_{\mathbf{s}^*_U} W^*_{\mathbf{t}}(\mathbf{s}^*_U) X^*(\mathbf{s}^*_U)] \\
&= 2^{-2k} \sum_{\mathbf{t}^*_W} \sum_{\mathbf{s}^*_U} W^*_{\mathbf{t}}(\mathbf{t}^*_W) W^*_{\mathbf{t}}(\mathbf{s}^*_U) E[X^*(\mathbf{t}^*_W) X^*(\mathbf{s}^*_U)] \\
&= 2^{-k} \sum_{\mathbf{t}^*_W} W^*_{\mathbf{t}}(\mathbf{t}^*_W) E[X^*(\mathbf{1}^*) X^*(\mathbf{t}^*_W)] \\
&= 2^{-k} \sum_{\mathbf{t}^*_W} W^*_{\mathbf{t}}(\mathbf{t}^*_W) \mathrm{v}_W = 2^{-k} \sum_{\mathbf{t}^*_W} W(\mathbf{t}) \mathrm{v}_W \\
&= 2^{-k} R(\mathbf{1}, \mathbf{t}),
\end{aligned}$$

since $W^*_{\mathbf{t}}(\mathbf{t}^*_W) = W(\mathbf{t})$.

Thus covariances for the process and interaction variances swap roles for X and X^*.

Example 2, continued. Construct the dual X^* of X by setting $\mathrm{var}(X^*(\mathbf{t}^*)) = \mathrm{v}_\emptyset$, $R^*_1 = R^*((1,1),(-1,1)) = \mathrm{v}_1$, $R^*_2 = R^*((1,1),(1,-1)) = \mathrm{v}_2$, and $R^*_{12} = R^*((1,1),(-1,-1)) = \mathrm{v}_{12}$. Then one can calculate the interaction variances for the dual process. For example, since $\mathrm{var}(\Gamma_{1^*}(X^*)) = \mathrm{v}_{1^*}$ is $\frac{1}{4}(1 - R^*_1 + R^*_2 - R^*_{12}) = \frac{1}{4}(\mathrm{v}_\emptyset - \mathrm{v}_1 + \mathrm{v}_2 - \mathrm{v}_{12})$, then $4\mathrm{v}_{1^*} = \frac{1}{4}(1 + R_1 + R_2 + R_{12}) - \frac{1}{4}(1 - R_1 + R_2 - R_{12}) + \frac{1}{4}(1 + R_1 - R_2 - R_{12}) - \frac{1}{4}(1 - R_1 - R_2 + R_{12}) = R_1$, and so forth.□

CHAPTER 3

Special Families of Stationary Processes

Certain stationary processes on T have properties that make them especially interesting. Section 3.1 defines orderly processes, a sensible class of processes in the context of experimental design. Another important family is isotropic processes, stationary processes represent equal prior knowledge for all factors. Correlations for an isotropic process depend only on distance and interaction variances depend only on word length. Some stationary processes that are particularly well-behaved are a special case of orderly isotropic processes, and are called tame processes. The Venn diagram in Figure 3.1 shows the relationship among these families.

After the present discussion of orderly, isotropic, and tame processes, Chapter 4 establishes certain closure properties for each family.

3.1 Orderly Processes

For the question of choosing an experimental design, we wish to consider processes with a realistic correspondence to an experimenter's prior knowledge of the factors. In analyzing the data from a multifactorial experiment, researchers standardly assume higher order effects are more likely to be negligible. Translating this assumption to the Bayesian formulation, a prior distribution for X should assign less variability to higher order interactions. Processes that incorporate

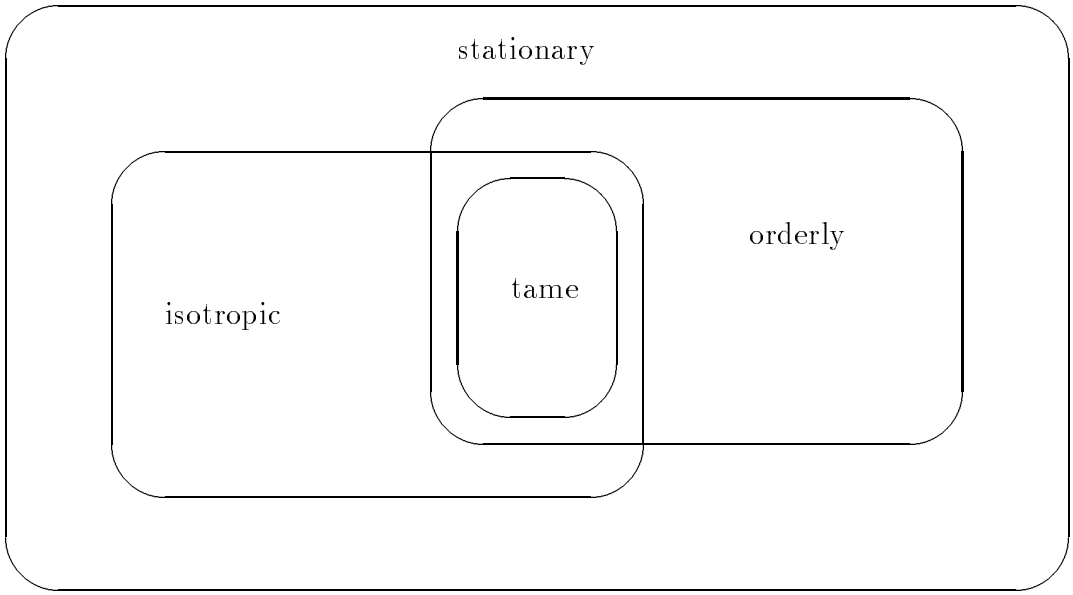

Figure 3.1: Families of Processes on T

this assumption have the property that $W \subset U \implies \mathrm{v}_W > \mathrm{v}_U$, 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. Stationary processes with both of these properties are called *orderly* processes.

It should be noted that nested decreasing interaction variances do not guarantee nested decreasing interaction variances, and vice versa. Section 3.2 addresses this issue in detail for isotropic processes.

When we say X is orderly, we implicitly assume the relevant conditions for any subprocesses. For instance, in (2.1) we assume Y and $\{\sum_W W(\mathbf{t})\beta_W\}$ are orderly processes. In particular, the β_W should have the property that $W \subset U \implies \mathrm{v}_W \geq \mathrm{v}_U$. Since error processes have constant interaction variances and independence between runs, adding an error process to and orderly process X preserves orderly-ness for X_i (as in (2.2)) without further assumptions.

3.2 Isotropic Processes

An *isotropic* process is a stationary process where the correlations depend only on distance, i.e. $EX(\mathbf{t})X(\mathbf{s})$ depends only on $d(\mathbf{t},\mathbf{s})$. 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). When discussing isotropic X we implicitly assume Y is isotropic and, in the setup with global effects, $\nu_W = \nu_{|W|}$. Note that an error

process ϵ is always isotropic so, for fixed i, X_i in (2.2) is isotropic whenever X is.

A stationary process X on T has 2^k parameters. The distribution of X is specified by 2^k covariances $R(\mathbf{1}, \mathbf{t})$ or, alternatively, by 2^k interaction variances v_W. For an isotropic process, $R(\mathbf{1}, \mathbf{t}) = R(d(\mathbf{1}, \mathbf{t}))$, and $\mathrm{v}_W = \mathrm{v}_{|W|}$, reducing the number of parameters to $k + 1$. This reduction is substantial for large k. In the context of experimental design, an isotropic process is appropriate when an experimenter has equal prior knowledge of all k experimental factors.

For an isotropic process, let $\mathbf{r}$ denote the vector of covariances $(r_0, r_1, \ldots, r_k)'$ and $\mathbf{v}$ denote the vector of interaction variances $(\mathrm{v}_0, \mathrm{v}_1, \ldots, \mathrm{v}_k)'$. From the conclusion of Section 2.3.2, $2^k\mathbf{v}$ provides the eigenvalues of the correlation matrix R for X on T. Therefore, it is precisely the positivity of $\mathbf{v}$ that determines which $\mathbf{r}$ legitimately specify an isotropic process. Section 3.2.1 establishes the relationship between $\mathbf{r}$ and $\mathbf{v}$.

3.2.1 Correlations, Interaction Variances, and Krawtchouk Polynomials

The linear transformation that takes $\mathbf{r}$ to $\mathbf{v}$ is given by Mitchell, Morris, and Ylvisaker ((1995), (2.4)):

$$\mathrm{v}_i = \frac{1}{2^k} \sum_{j=0}^{k} r_j \sum_{m} \binom{i}{m} \binom{k-i}{j-m} (-1)^m.$$

Similarly, one can find the linear transformation that takes $\mathbf{v}$ to $\mathbf{r}$. Let $\mathbf{t}_i$ be the run in T with -1 in the first i positions, 1 in the last $k - i$ positions.

$$\begin{aligned} r_i &= \mathrm{cov}(X(\mathbf{1}), X(\mathbf{t}_i)) = \mathrm{cov}(\sum_W \Gamma_W, \sum_U U(\mathbf{t}_i)\Gamma_U) = \sum_W W(\mathbf{t}_i)\mathrm{v}_{|W|} \\ &= \sum_{j=0}^{k} \mathrm{v}_j \sum_{W:|W|=j} W(\mathbf{t}_i) = \sum_{j=0}^{k} \mathrm{v}_j \sum_{m} \binom{i}{m} \binom{k-i}{j-m} (-1)^m \end{aligned}$$

This is the same transformation aside from a factor of 2^k. In matrix notation write $\mathbf{v} = 2^{-k} K\mathbf{r}$ and $\mathbf{r} = K\mathbf{v}$. The entries in the matrix K are easily described by employing Krawtchouk polynomials.

For a fixed positive integer k, the j^{th} *Krawtchouk polynomial* $P_j(x;k)$ is defined by

$$P_j(x;k) = \sum_m (-1)^m \binom{x}{m}\binom{k-x}{j-m}, \quad j = 0, 1, \ldots, k.$$

So for the matrix $K = K_k$ defined above, $K(i,j) = P_j(i;k)$. The transformations between $\mathbf{r}$ and $\mathbf{v}$ can be written $r_i = \sum_j P_j(i;k)\mathrm{v}_j$ and $\mathrm{v}_i = 2^{-k}\sum_j P_j(i;k) r_j$.

Example 3 Let X be a stationary process with isotropic correlations given by $R(\mathbf{s},\mathbf{t}) = \rho^{d(\mathbf{s},\mathbf{t})}$. Then $\mathrm{v}_i = 2^{-k}\sum_j P_j(i;k)\rho^j$. Krawtchouk polynomials have the generating function $(1+z)^{k-i}(1-z)^i = \sum_{j=0}^k P_j(i;k) z^j$, so the i^{th} interaction variance is $2^{-k}(1+\rho)^{k-i}(1-\rho)^i$. Note that X is orderly as well as isotropic. In fact, X is a particular kind of isotropic orderly process called a tame process, defined in the next section. □

3.3 Tame Processes

Section 3.1 discussed orderly processes, which have nested decreasing interaction variances and nested decreasing covariances. For an isotropic process, nested decreasing interaction variances reduces to the condition that $\mathrm{v}_i > \mathrm{v}_{i+1}$ for $i = 0, 1, \ldots, k-1$. Nested decreasing covariances means $r_i > r_{i+1}$ for $i = 0, 1, \ldots, k-1$. As noted by Mitchell, Morris, and Ylvisaker (1995), it is not enough to assume the v_i decrease to guarantee the r_i will as well. 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 operations D^m

for a finite sequence a_j, $j = 0, 1, \ldots, k$, as follows:

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

Call the finite sequence a_j, $j = 0, 1, \ldots, k$, *completely monotone to order* l provided $D^m a_j \geq 0$ for all $j + m \leq l$, where $l \leq k$. If $l = k$ then simply say the sequence a_j is completely monotone.

Proposition 3.3.1 *If the interaction variances* v_j, $j = 0, 1, \ldots, k$, *are completely monotone, then the associated isotropic process has completely monotone correlations* r_i, $i = 0, 1, \ldots, k$. *Similarly, an isotropic process with completely monotone correlations* r_i *has completely monotone interaction variances* v_j.

The proof of Proposition 3.3.1 is deferred to an appendix.

Proposition 3.3.1 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, since they are "balanced" across factors (isotropic) and their correlations and interaction variances decrease "smoothly" (completely monotone). Yet for design purposes, an experimenter with roughly equal prior knowledge of his or her experimental factors will likely find tame processes general enough to include any isotropic process he or she might consider.

Remark 1. It is known that any infinite sequence a_i, $i = 0, 1, \ldots,$ is completely monotone if and only if $a_n = \int \rho^n dG(\rho)$ for some distribution G on the unit interval (see e.g. Theorem 6 in Widder (1971)). However, not every finite completely monotone sequence can be extended to an infinite completely monotone sequence. In the case $k = 2$, take $a_0 = 1, a_1 = 1/2, a_2 = 0$, for example. The only possibility for appending a_3 and keeping the sequence non-negative and non-increasing is $a_3 = 0$. But then the third difference D^3 is $1 - 3(\frac{1}{2}) + 3(0) - 0 < 0$.

This is an example of a finite completely monotone sequence that cannot be extended to an infinite completely monotone sequence, and thus does not arise as $\int \rho^n dG(\rho)$. Therefore the representation $a_n = \int \rho^n dG(\rho)$ yields some, but not all, finite completely monotone sequences.

Remark 2. The set of tame processes is in fact smaller than the set of orderly isotropic processes. Let $k = 4$ and assume the r_i and v_i are both positive and decreasing (in other words, correspond to an orderly isotropic process). These assumptions and the Krawtchouk relationship between the two sequences are enough to imply that most of the applicable difference operators applied to these sequences must be non-negative. In fact, every applicable difference operator must be non-negative except $D^2 r_2$ and $D^2 \mathrm{v}_2$. Consider the isotropic correlations $\mathbf{r} = (1, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}, 0)'$ with interaction variances $\mathbf{v}$ proportional to $(4, \frac{4}{3}, \frac{2}{3}, \frac{2}{3}, 0)'$. (One could perturb the values slightly so that the sequences are strictly positive and strictly decreasing.) This represents an orderly isotropic process that is not a tame process.

3.3.1 Tame Processes with Global Effects

In the more general setup (2.1), when we say the observable process X is tame, we implicitly assume Y is as well. However, added care is needed with respect to the global effects β_W. The problem is that with $\nu_W = 0$ for $|W| > L$, the process $\sum_W W(\mathbf{t})\beta_W$ is not necessarily tame.

In general, if one includes global effects for interactions of degree up to L, so

$$X(\mathbf{t}) = \beta_O + \sum_i t_i \beta_i + \sum_{i \neq j} t_i t_j \beta_{ij} + \ldots + \sum_{W:|W|=L} W(\mathbf{t})\beta_W + Y(\mathbf{t}) + \epsilon(\mathbf{t}), \quad (3.1)$$

then

$$r_i^X = \nu_0 + P_1(i;k)\nu_1 + P_2(i;k)\nu_2 + \ldots + P_L(i;k)\nu_L + r_i^Y + 2^{-k}\sigma^2. \quad (3.2)$$

(The case $\sigma^2 = 0$ is included.) When r_i^Y is completely monotone, certain conditions on the ν_i are required so that r_i^X is as well.

Viewing $P_j(i;k)$ as a sequence in i for $i = 0, 1, \ldots, k$, the following lemma helps describe conditions on the ν_i necessary to guarantee complete monotoneity for r_i^X.

Lemma 3.3.1 *Let $m + i \leq k$. If $m > j$, $D^m P_j(i;k) = 0$. In addition (for the case $m = j$), $D^m P_m(i;k) = 2^m > 0$.*

Proof: $D^m P_j(i;k) = \sum_{h=0}^m \binom{m}{h}(-1)^h P_j(i+h;k)$ and $P_j(i+h;k)$ is a polynomial in $i+h$ of degree j, and thus a polynomial in h of degree j. Since $\sum_{h=0}^m \binom{m}{h}(-1)^h h^j = 0$ for $m > j$, the first statement in the lemma follows. To prove the second statement, verify that the coefficient for h^m in $P_m(i+h;k)$ is $(-2)^m/m!$. Then $D^m P_m(i;k) = \sum_{h=0}^m \binom{m}{h}(-1)^h P_m(i+h;k) = [(-2)^m/m!] \sum_{h=0}^m \binom{m}{h}(-1)^h h^m$ since all smaller powers of h vanish. Recall that $\sum_{h=0}^m \binom{m}{h}(-1)^h h^m = (-1)^m m!$ and the result follows. □

Clearly there is no problem adding just a global mean, because then $r_i^X = \nu_0 + r_i^Y + 2^{-k}\sigma^2$ and complete monotoneity of r_i^Y guarantees the same for r_i^X. But consider adding a global mean and main effects, so that $X(\mathbf{t}) = \beta_O + \sum_i t_i \beta_i + Y(\mathbf{t}) + \epsilon(\mathbf{t})$ and (3.2) becomes $r_i^X = \nu_0 + (k-2i)\nu_1 + r_i^Y + 2^{-k}\sigma^2$. Because of Lemma 3.3.1, the only concern is whether $D^0 r_i^X \geq 0$, which is to say that the r_i^X are non-negative. For this one must have $\nu_0 + (k-2i)\nu_1 + r_i^Y + 2^{-k}\sigma^2 \geq 0$ for all i, or $(\nu_o + r_i^Y + 2^{-k}\sigma^2)/(2i-k) \geq \nu_1$ for all $i \geq k/2$. It is necessary and sufficient that $\nu_1 \leq (\nu_o + r_k^Y + 2^{-k}\sigma^2)/k$, and sufficient that $k\nu_1 \leq \nu_0$. With this condition, complete monotoneity is preserved for r_i^X

Increasingly complicated conditions on the relative sizes of the ν_i arise as L increases. One way to avoid the problem altogether is to discard the assumption

that the ν_W vanish after some threshold and require $\nu_{|W|}$ be completely monotone.

CHAPTER 4

Closure Properties for Families of Processes

It turns out that families of processes defined and discussed in the previous chapter have pleasant closure properties. We identify and establish these properties here. The properties resurface in Chapter 8 when we concentrate on design questions.

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 in the family with covariances R_1, $R_2 \in \mathcal{R}$ and with interaction variances V_1, V_2. We will say X_i is distributed as R_i, for short. (Throughout this chapter, subscripts on processes are used to differentiate them, and not to suggest replications of a process observed with error as at (2.2)). Desirable properties for $\mathcal{R}$ are:

(P1) Closure under positive linear combinations:

$R_1, R_2 \in \mathcal{R}$, $a, b > 0 \implies aR_1 + bR_2 \in \mathcal{R}$

(P2) Closure under direct products: R_1, $R_2 \in \mathcal{R} \implies R_1 \otimes R_2 \in \mathcal{R}$

(P3) Closure under matrix products: R_1, $R_2 \in \mathcal{R} \implies R_1 R_2 \in \mathcal{R}$

Note that convexity is a special case of (P1).

We first verify that the set of stationary processes has these three properties, then show the families of orderly, isotropic, and tame processes do as well.

4.1 Stationary Processes

Let X_1 and X_2 be independent stationary processes distributed as R_1 and R_2 and let $a, b > 0$. Then the process $X = \sqrt{a}X_1 + \sqrt{b}X_2$ is Gaussian and has distribution $aR_1 + bR_2$. Since $E[X(\mathbf{s})X(\mathbf{t})] = aR_1(\mathbf{s},\mathbf{t}) + bR_2(\mathbf{s},\mathbf{t}) = aR_1(\mathbf{su},\mathbf{tu}) + bR_2(\mathbf{su},\mathbf{tu}) = E[X(\mathbf{su})X(\mathbf{tu})]$, X is stationary. Thus property (P1) holds.

Now let X be a Gaussian process with the same covariances as the process X_1X_2. Then

$$\begin{aligned} E[X(\mathbf{s})X(\mathbf{t})] &= E[X_1(\mathbf{s})X_1(\mathbf{t})]E[X_2(\mathbf{s})X_2(\mathbf{t})] \\ &= E[X_1(\mathbf{su})X_1(\mathbf{tu})]E[X_2(\mathbf{su})X_2(\mathbf{tu})] \\ &= E[X(\mathbf{su})X(\mathbf{tu})]. \end{aligned}$$

So X is again stationary, and has covariance $R_1 \otimes R_2$. Thus stationary processes are closed under direct products, property (P2).

Matrix products of stationary covariance matrices R_i are easily gotten using $R_i = 2^{2k}O'V_iO$, where $O'O = 2^{-k}I$ and V_i is the diagonal matrix of interaction variances for R_i. Because O is orthogonal, we have $R_1R_2 = 2^{3k}O'V_1V_2O$. This shows R_1R_2 can serve as the covariance of a stationary process because the interaction variances are independent. Stationary processes are closed under matrix products, property (P3).

4.2 Orderly Processes

Recall that orderly processes have positive covariances such that $R(\mathbf{t},\mathbf{s}) > R(\mathbf{t},\mathbf{u})$ whenever $\mathbf{s}$ is between $\mathbf{t}$ and $\mathbf{u}$, and interaction variances such that $W \subset U \implies \mathrm{v}_W > \mathrm{v}_U$. It is easy to check that these properties are preserved by taking positive linear combinations. Properties (P2) and (P3) require a bit more

effort. Property (P2) is an immediate consequence of the following proposition, and (P3) follows in a corollary.

Proposition 4.2.1 *Let X_1 and X_2 be orderly processes, and let X be a Gaussian process with the covariance of X_1X_2. Then X is orderly.*

We defer the proof of Proposition 4.2.1 to an appendix.

Corollary 4.2.1 *Orderly processes have property (P3).*

Proof: Recall that $R_1R_2 = 2^{3k}O'V_1V_2O$, so it suffices to show that V_1V_2 is the matrix of interaction variances for some orderly process.

Note that a process is orderly if and only if its dual process is orderly. Let X_1^* and X_2^* be the dual processes for X_1 and X_2. Then $X^* = X_1^*X_2^*$ is orderly by Proposition 4.2.1.

Now X_1^* and X_2^* have covariances given by V_1 and V_2, respectively, and X^* has covariances given by V_1V_2. Since X^* is orderly, so is its dual. But the dual of X^* has interaction variances proportional to V_1V_2, the required result. □

4.3 Isotropic Processes

If R_1 and R_2 are isotropic covariance matrices, clearly so are $aR_1 + bR_2$ and $R_1 \otimes R_2$. Recall that a process is isotropic if and only if its interaction variances v_W depend only on $|W|$. But this property is preserved by taking products of v_W. Since $R_1R_2 = 2^{3k}O'V_1V_2O$, R_1R_2 has interaction variances proportional to such products and so also represents an isotropic process.

4.4 Tame Processes

By the linearity of the difference operators, the set of covariance matrices for tame processes on T is closed under addition and multiplication by a positive scalar. and so has property (P1). Properties (P2) and (P3) are corollaries of the following proposition, whose proof is deferred to an appendix.

Lemma 4.4.1 *The set of completely monotone $k+1$-vectors is closed under direct multiplication.*

Corollary 4.4.2 *The set of tame processes on T has properties (P2) and (P3).*

For two tame processes X_1 and X_2, with covariances $r_{1,j}$ and $r_{2,j}$, the product process $X = X_1 X_2$ has covariances $r_{1,j} r_{2,j}$, which form a completely monotone sequence by Lemma 4.4.1. Therefore X is tame by Proposition 3.3.1. Similarly, Lemma 4.4.1 implies that $\mathrm{v}_{1,j}\mathrm{v}_{2,j}$ are completely monotone whenever $\mathrm{v}_{1,j}$ and $\mathrm{v}_{2,j}$ are, which implies that the process with interaction variances $\mathrm{v}_{1,j}\mathrm{v}_{2,j}$ is completely monotone by Proposition 3.3.1. □

CHAPTER 5

Stationary Processes and Subgroups of T

In Chapters 3 and 4 we described families of stationary distributions appropriate for describing an experimenter's prior knowledge of his or her factors. We now turn to studying the covariance structure of stationary processes.

The covariance matrix of a stationary process X on T has a particular structure, and the interactions $\Gamma_W(X)$ are a useful mathematical tool to describe it. As described in Section 2.3.2, 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 with the interaction variances along the diagonal. 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, namely those corresponding to X restricted to a subgroup of T. This is explained in Section 5.1. Using these results, Section 5.2 derives the posterior distribution of X conditioned on a subgroup of T. These posterior variances and covariances become important in Chapter 6 when discussing design criteria, and thus move us closer to addressing the problem of choosing a design.

5.1 Covariance of Stationary Processes on Subgroups of T

The definition of stationarity is essentially invariance under group translation of runs. However, subgroups of T are necessarily closed under multiplication. Let F be a subgroup of T. Considering X as a process on F, for any given $\mathbf{f}$, $Z(\mathbf{t}) = X(\mathbf{ft})$ is well-defined as a process on F because $\mathbf{f}, \mathbf{t} \in F \implies \mathbf{ft} \in F$. 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.

This suggests that the structure of the covariance matrix for X on T described in Section 2.3.2 might generalize to covariance matrices of such restricted processes. This section shows this to be the case. The approach taken here is an alternative to that used in Section 3 of Mitchell, Morris, and Ylvisaker (1995).

For the process X restricted to a subgroup F, write R_F for the corresponding covariance matrix.

Viewing T as a vector space, F is a subspace. As such, F has a perpendicular subspace $F^{\perp}$. Now T is isomorphic to the group of words $\mathcal{W}$. Map the runs in $F^{\perp}$ to $\mathcal{W}$ via the natural isomorphism and denote the image as $\mathcal{A}_F$. Then $\mathcal{A}_F$ is a subgroup of $\mathcal{W}$. By its definition, $\mathcal{A}_F = \{W \in \mathcal{W} : W(\mathbf{f}) = 1 \ \forall \ \mathbf{f} \in F\}$. Say $|\mathcal{A}_F| = 2^p$ for some $0 \leq p \leq k$, so that $|F| = 2^{k-p}$.

Call the cosets of $\mathcal{A}_F$ *alias sets* and let $\mathcal{A}$ denote a generic coset. There are 2^{k-p} alias sets, each with 2^p words, partitioning the word group. Note that this partition of words induces a partition of interactions, so one can naturally speak of alias sets of interactions without confusion.

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, this definition immediately shows alias interactions are independent because independence is inherited from the Γ_W. Moreover, we will show alias interactions play the same role for X restricted to F as interactions play for X on T. To that end, our first task is to show that each alias interaction depends only on $X(\mathbf{f})$ for $\mathbf{f}$ in F.

Proposition 5.1.1 *For a subgroup F of T with alias set $\mathcal{A}$, the alias interaction $\Gamma_{\mathcal{A}}$ can be written $2^{-k}\sum_{\mathbf{f}\in F}\sum_{W\in\mathcal{A}} X(\mathbf{f})W(\mathbf{f})$. In particular, any alias interaction depends only on $X(\mathbf{f})$ for $\mathbf{f}\in F$.*

The proof of Proposition 5.1.1 is deferred to an appendix.

We can simplify $2^{-k}\sum_{\mathbf{f}\in F}\sum_{W\in\mathcal{A}} X(\mathbf{f})W(\mathbf{f})$ to derive a more pleasant representation for alias interactions. Observe that for fixed $\mathbf{f}\in F$, W and U in the same alias set guarantees $W(\mathbf{f}) = U(\mathbf{f})$. This follows immediately for $\mathcal{A}_F$ and then for any $\mathcal{A}$ by translation of $\mathcal{A}_F$. So for $\mathbf{f}\in F$, the notation $\mathcal{A}(\mathbf{f})$ is well-defined. A convenient expression for an alias interaction is then

$$\Gamma_{\mathcal{A}} = 2^{-k}\sum_{\mathbf{f}\in F} X(\mathbf{f})\sum_{W\in\mathcal{A}} W(\mathbf{f}) = 2^{-(k-p)}\sum_{\mathbf{f}\in F}\mathcal{A}(\mathbf{f})X(\mathbf{f}). \tag{5.1}$$

Just as a function on T can be recovered from its interactions, the process X on F can be recovered from its alias interactions. If $\mathbf{f}$ is in F, then

$$\begin{aligned}\sum_{\mathcal{A}}\mathcal{A}(\mathbf{f})\Gamma_{\mathcal{A}} &= \sum_{\mathcal{A}}\mathcal{A}(\mathbf{f})\sum_{W\in\mathcal{A}}\Gamma_W = \sum_{\mathcal{A}}\sum_{W\in\mathcal{A}} W(\mathbf{f})\Gamma_W \\ &= \sum_{W} W(\mathbf{f})\Gamma_W = X(\mathbf{f}).\end{aligned} \tag{5.2}$$

Now rewrite the transformation in (5.1) in matrix notation. Let $\Gamma_{\mathcal{A}}$ be the vector of the 2^{k-p} alias interactions, $\mathbf{X}_F$ be the vector of $X(\mathbf{f})$ for $\mathbf{f}\in F$, and O_F be the $2^{k-p}\times 2^{k-p}$ matrix that transforms $\mathbf{X}_F$ to $\Gamma_{\mathcal{A}}$. Then $\Gamma_{\mathcal{A}} = O_F\mathbf{X}_F$, and

so $V_F = \mathrm{var}(\Gamma_{\mathcal{A}}) = O_F R_F O_F'$. Since the alias interactions are independent, V_F is diagonal with the variances of the alias interactions along the diagonal. The variance of an alias interactions is the sum of the variances of the interactions in an alias set.

Just as O was an orthogonal transformation of X on T, O_F is an orthogonal transformation of X on F. To see this, first recall that for the alias set $\mathcal{A}_F$, $\mathcal{A}_F(\mathbf{f}) = 1$ for all $\mathbf{f}$ in F, by definition. Any other alias set is a coset of $\mathcal{A}_F$ and can be written as $W\mathcal{A}_F$ for some $W \notin \mathcal{A}_F$. Since $W \notin \mathcal{A}_F$, there exists a run $\mathbf{f}$ in F such that $W(\mathbf{f}) = -1$. But $\mathbf{1}$ is in F and $W(\mathbf{1}) = 1$. One can set up a bijection (as in the proof of Lemma D.0.1) to show that W (and, equivalently, $W\mathcal{A}_F$) is "1" on exactly half the elements of F and "-1" on the other half. Moreover, two different alias sets $W\mathcal{A}_F$ and $U\mathcal{A}_F$ agree on exactly half the elements of F. Therefore O_F is orthogonal.

Since $2^{(k-p)/2}O_F$ is orthonormal and $V_F = O_F R_F O_F'$, we can factor R_F as $R_F = (2^{(k-p)/2}O_F')2^{k-p}V_F(2^{(k-p)/2}O_F)$. As noted, V_F is diagonal because alias interactions are independent, 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.

The next section shows this decomposition of R_F is useful in considering processes conditioned on subgroups.

5.2 Processes Conditioned on Subgroups and Posterior Distributions

Start with a stationary process X on T and consider the process conditioned on a subgroup F of T. In other words, suppose one has observations (with or without error) of $X(\mathbf{f})$ for $\mathbf{f}$ in F and is left with a process on the remaining runs

in T. The decomposition of R_F derived in the previous section aids in deriving posterior variances and covariances for the conditioned process. Moreover, the conditioned process retains certain limited stationary properties.

For notational ease, we sometimes use σ_p^2 to denote $2^{-(k-p)}\sigma^2$ and var_F and cov_F to denote posterior variances and covariances for a process conditioned on a subgroup F.

As in Mitchell, Morris, and Ylvisaker (1995), Proposition 3.3, consider the covariance of an ordinary interaction with a single observation.

$$\mathrm{cov}(\Gamma_W, X(\mathbf{f})) = E[\Gamma_W \sum_U \Gamma_U U(\mathbf{f})] = E[\Gamma_W^2 W(\mathbf{f})] = W(\mathbf{f})\mathrm{v}_W \tag{5.3}$$

Compare (5.3) with (5.1) to find that the vector of covariances of Γ_W with the process X on F is $2^{k-p}\mathrm{v}_W$ times the row of O_F that corresponds to $\Gamma_{W\mathcal{A}_F}$. (Note that $W\mathcal{A}_F$ is the alias set containing W.) The orthogonality of O_F, the decomposition $R_F = 2^{2(k-p)}O_F' V_F O_F$, and the fact that $\sigma^2 I$ can be written $\sigma^2 2^{k-p} O_F' O_F$ imply

$$\mathrm{var}(\Gamma_W | X(\mathbf{f}) + \epsilon(\mathbf{f}), \mathbf{f} \in F) = \mathrm{v}_W - \frac{\mathrm{v}_W^2}{\mathrm{v}_{(W\mathcal{A}_F)} + \sigma_p^2}, \tag{5.4}$$

$$\mathrm{cov}(\Gamma_W, \Gamma_U | X(\mathbf{f}) + \epsilon(\mathbf{f}), \mathbf{f} \in F) = -\frac{\mathrm{v}_W \mathrm{v}_U}{\mathrm{v}_{(W\mathcal{A}_F)} + \sigma_p^2}\delta_{(W\mathcal{A}_F)(U\mathcal{A}_F)}. \tag{5.5}$$

(The case of error-free observation is included by setting $\sigma^2 = 0$). Note that (5.5) implies that interactions in different alias sets remain independent in the conditioned process, but interactions that are aliased by F become dependent.

Obviously the conditioned process is no longer stationary since interactions aliased by F are dependent in the posterior process. However, certain limited stationarity properties remain. These properties hold for cosets of F.

Consider posterior covariances:

$$\begin{aligned}\mathrm{cov}_F(X(\mathbf{s}), X(\mathbf{t})) &= \mathrm{cov}_F\left(\sum_W W(\mathbf{s})\Gamma_W, \sum_U U(\mathbf{t})\Gamma_U\right) \\ &= \sum_W \sum_U W(\mathbf{s})U(\mathbf{t})\mathrm{cov}_F(\Gamma_W, \Gamma_U)\end{aligned}$$

The posterior covariance of Γ_W and Γ_U remains 0 unless F aliases W and U. But whenever W and U are aliased, WU is in $\mathcal{A}_F$. So then for any $\mathbf{f} \in F$, $WU(\mathbf{f}) = 1$ and

$$\begin{aligned}\mathrm{cov}_F(X(\mathbf{s}), X(\mathbf{t})) &= \sum_W \sum_{U \in W\mathcal{A}_F} W(\mathbf{s})U(\mathbf{t})\mathrm{cov}_F(\Gamma_W, \Gamma_U)WU(\mathbf{f}) \\ &= \sum_W \sum_{U \in W\mathcal{A}_F} W(\mathbf{sf})U(\mathbf{tf})\mathrm{cov}_F(\Gamma_W, \Gamma_U) = \mathrm{cov}_F(X(\mathbf{sf}), X(\mathbf{tf}))\end{aligned}$$

Writing X_F for the conditioned process, $\{X_F(\mathbf{t})\}$ has the same covariance structure as $\{Z_F(\mathbf{t})\}$, where $Z_F(\mathbf{t}) = X_F(\mathbf{tf}_0)$ for some fixed $\mathbf{f}_0 \in F$. (One cannot say $\{X_F(\mathbf{t})\}$ is distributed as $\{Z_F(\mathbf{t})\}$ since X_F and Z_F may have different means.) Multiplying any $\mathbf{t} \in T$ by an element of F translates $\mathbf{t}$ within its coset. This gives stationarity (ignoring means) for the posterior process on any coset of F. In particular, all runs in the same coset of F have the same posterior variance.

Finally, since $X(\mathbf{t})$ can be recovered from interactions, one can use (5.4) and (5.5) to derive posterior variances for unobserved runs, although the expression

turns out to be more complicated and not particularly instructive.

$$\begin{aligned}
\mathrm{var}_F(X(\mathbf{t})) &= \mathrm{var}_F(\sum_W W(\mathbf{t})\Gamma_W) = \sum_{\mathcal{A}} \mathrm{var}_F(\sum_{W\in\mathcal{A}} W(\mathbf{t})\Gamma_W) \\
&= \sum_{\mathcal{A}} \left[\sum_{W\in\mathcal{A}} (\mathrm{v}_W - \frac{\mathrm{v}_W^2}{\mathrm{v}_{\mathcal{A}}+\sigma_p^2}) + 2 \sum_{W\neq U\in\mathcal{A}} W(\mathbf{t})U(\mathbf{t}) \frac{-\mathrm{v}_W \mathrm{v}_U}{\mathrm{v}_{\mathcal{A}}+\sigma_p^2} \right] \\
&= \sum_{\mathcal{A}} \left[\mathrm{v}_{\mathcal{A}} - \frac{(\sum_{W\in\mathcal{A}} W(\mathbf{t})\mathrm{v}_W)^2}{\mathrm{v}_{\mathcal{A}}+\sigma_p^2} \right] \qquad (5.6) \\
&= \sum_{\mathcal{A}} \left[\frac{\mathrm{v}_{\mathcal{A}}\sigma_p^2}{\mathrm{v}_{\mathcal{A}}+\sigma_p^2} + \frac{4}{\mathrm{v}_{\mathcal{A}}+\sigma_p^2} \sum_{\substack{W,U\in\mathcal{A}: \\ W(\mathbf{t})\neq U(\mathbf{t})}} \mathrm{v}_W \mathrm{v}_U \right] \qquad (5.7)
\end{aligned}$$

Notice that for all runs in F, the second part of each summand in (5.7) vanishes since all words in the same alias set agree (as functions) on design sites. In addition, the terms $\mathrm{v}_{\mathcal{A}}\sigma_p^2/(\mathrm{v}_{\mathcal{A}}+\sigma_p^2)$ vanish when there is no observational error.

CHAPTER 6

Stationary Processes and Design

An underlying theme for this work is that stationary processes on T provide a framework for thinking about experimentation with multiple two-level factors. The design space is T, and we imagine observations as realizations of a stochastic process $X(\mathbf{t})$ or, in the presence of experimental error, $X(\mathbf{t}) + \epsilon(\mathbf{t})$. The stationarity assumption for X ensures a kind of "impartiality" in the expected distribution of observations.

Given insufficient resources to observe all 2^k runs in T, which subset of T should one choose for observation? We consider choosing among the classical fractional factorial designs. The results of Chapter 5 immediately pertain, since a fractional factorial is nothing more than a subgroup of T.

After establishing some terminology and notation for fractional factorials in Section 6.1, we turn to a discussion of various optimality criteria. Up to this point, interactions have been used as a mathematical tool as a transform of a stationary process. However, interactions by themselves are interesting to researchers because they correspond to factor "effects." Mathematically and practically, for some criteria it makes sense to discuss optimality for interactions as much as for the process itself.

6.1 Fractional Factorials

A regular *fractional factorial* design for k two-level factors is usually specified by p independent words and the "defining relation" $\emptyset = W_1 = W_2 = \cdots = W_p$. The set of runs in the fractional factorial is then $F = \{\mathbf{t} \in T | W_i(\mathbf{t}) = 1 \ \forall \ i = 1, ..., p\}$. Note that F is a subgroup of T. The words $W_1, W_2, \cdots, W_p$ span a subgroup of $\mathcal{W}$. This subgroup of $\mathcal{W}$ is the subgroup $\mathcal{A}_F$ defined in Section 5.1. The subgroup $\mathcal{A}_F$ is sometimes called the set of defining relations. As before, we refer to cosets of $\mathcal{A}_F$ as alias sets and denote a generic alias set as $\mathcal{A}$.

A few more bits of notation for fractional factorials are useful in later chapters. For a fractional factorial F define

$$D_i = |\{\mathbf{f} \in F \ : \ d(\mathbf{1}, \mathbf{f}) = i\}|,$$
$$L_i = |\{W \in \mathcal{A}_F \ : \ |W| = i\}|.$$

The vectors $(D_0, D_1, \ldots, D_k)'$ and $(L_0, L_1, \ldots, L_k)'$ are the *design distance* and *word length* vectors, respectively. Note that replacing $\mathbf{1}$ with any fixed run in F in the definition of the D_i yields the same design distance vector. 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 maximum resolution–minimum aberration, formulated by Fries and Hunter (1980).

It is known that Krawtchouk polynomials, introduced in Section 3.2.1, can be used to compute the word lengths from the design distances, and vice-versa (MacWilliams and Sloane (1977), p. 129):

$$D_i = \frac{1}{2^{k-p}} \sum_{j=0}^{k} P_i(j; k) L_j \quad \text{and} \quad L_i = \frac{1}{2^p} \sum_{j=0}^{k} P_i(j; k) D_j. \tag{6.1}$$

Example 4 For $k = 5$ factors consider the 2^{5-2} design with defining relation $\emptyset = 123 = 345$. Then $\mathcal{A}_F$=$\{\emptyset, 123, 345, 1245\}$. Since there are two words of length 3 and one word of length 4, the design distance vector $\mathbf{L}$ is $(1, 0, 0, 2, 1, 0)'$. The eight runs in the fractional factorial F are

$$
\begin{array}{ccccccc}
(& 1, & 1, & 1, & 1, & 1 &), \\
(& 1, & 1, & 1, & -1, & -1 &), \\
(& 1, & -1, & -1, & 1, & -1 &), \\
(& 1, & -1, & -1, & -1, & 1 &), \\
(& -1, & 1, & -1, & 1, & -1 &), \\
(& -1, & 1, & -1, & -1, & 1 &), \\
(& -1, & -1, & 1, & 1, & 1 &), \\
(& -1, & -1, & 1, & -1, & -1 &).
\end{array}
$$

Two runs are Hamming distance 2 from $\mathbf{1}$, four runs are Hamming distance 3, and one run is Hamming distance 4, so design distance vector $\mathbf{D}$ is $(1, 0, 2, 4, 1, 0)'$. The design has resolution 3 and aberration 2.□

As in Chapter 5, when X is restricted to a fractional factorial F, write R_F for the corresponding covariance matrix.

The remainder of this chapter defines and discusses some criteria for choosing among fractional factorials. Each criterion is translated into the mathematical notation of the prior distribution of X. We derive the optimality criteria in generality, including the possibility of observational error. For the case of error-free observation, set $\sigma^2 = 0$ (which implies $\sigma_p^2 = 0$).

6.2 D-Optimality

The D-optimal fractional factorial of a given size is the fractional factorial that minimizes the posterior generalized variance of the process X at unobserved runs in T. Since X is Gaussian, this posterior generalized variance equals $|R|/|R_F|$ in the case of error-free observation, so the D-optimal design maximizes $|R_F|$. Similarly, when $\sigma^2 > 0$ the D-optimal design maximizes $|R_F + \sigma^2 I|$ (Mitchell, Sacks, and Ylvisaker (1994)).

Since R_F and I are simultaneously diagonalizable by the orthonormal matrix $2^{(k-p)/2}O_F$, $|R_F + \sigma^2 I| = |2^{(k-p)/2}O_F'| \cdot |V_F + \sigma^2 I| \cdot |2^{(k-p)/2}O_F| =$

$$= \prod_{\mathcal{A}} (2^{k-p}\mathrm{v}_{\mathcal{A}} + \sigma^2) = (2^{k-p})^{2^{k-p}} \prod_{\mathcal{A}} (\sigma_p^2 + \sum_{W \in \mathcal{A}} \mathrm{v}_W). \tag{6.2}$$

When comparing fractions of the same size (constant p), ignore the power of 2 in (6.2). Thus the D-optimal fractional factorial of a given size partitions the word group to maximize

$$D(F) = \prod_{\mathcal{A}} (\sigma_p^2 + \sum_{W \in \mathcal{A}} \mathrm{v}_W). \tag{6.3}$$

6.3 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.

The interactions are an orthogonal transformation of the 2^k potential runs,

$\Gamma = O\mathbf{X}$. Recall that $R = \text{var}(\mathbf{X}) = (2^{k/2}O)'2^k V(2^{k/2}O)$ with $2^{k/2}O$ orthonormal and $V = \text{var}(\Gamma)$. This implies $\text{trace}(R) = \text{trace}(2^k V)$. 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, $\text{trace}[\text{var}(\mathbf{X})|X(\mathbf{f}) + \epsilon(\mathbf{f}), \mathbf{f} \in F] = \text{trace}[\text{var}(\Gamma)|X(\mathbf{f}) + \epsilon(\mathbf{f}), \mathbf{f} \in F)]$. This discussion proves Proposition 6.3.1.

Proposition 6.3.1 *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 6.3.1 allows us to sum the posterior variances of the interactions than the posterior variances of the $X(\mathbf{t})$. Referring to (5.4), the A-optimal fractional factorial minimizes

$$\sum_{\mathcal{A}} \sum_{W \in \mathcal{A}} (\text{v}_W - \frac{\text{v}_W^2}{\text{v}_{\mathcal{A}} + \sigma_p^2}).$$

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

$$A(F) = \sum_{\mathcal{A}} \frac{\sum_{W \in \mathcal{A}} \text{v}_W^2}{\sigma_p^2 + \sum_{W \in \mathcal{A}} \text{v}_W}. \tag{6.4}$$

6.4 G-Optimality

The G-optimal fractional factorial F of a given size is the fraction that minimizes $\max_{\mathbf{t}} \text{var}_F(X(\mathbf{t}))$. Refer to the expressions for $\text{var}_F(X(\mathbf{t}))$ at the end of Section 5.2. By the remark following (5.7), the largest posterior variance must occur at a run not in the design. Referring to (5.6), since $\sum_{\mathcal{A}} \text{v}_{\mathcal{A}}$ is constant over

designs, the G-optimal design maximizes

$$\min_{\mathbf{t}\notin F} \sum_{\mathcal{A}} \left[\frac{\sum_{W\in\mathcal{A}} \mathrm{v}_W^2}{\mathrm{v}_{\mathcal{A}} + \sigma_p^2} + 2\frac{\sum_{W\neq U\in\mathcal{A}} WU(\mathbf{t})\mathrm{v}_W\mathrm{v}_U}{\mathrm{v}_{\mathcal{A}} + \sigma_p^2} \right]. \tag{6.5}$$

Note the first part of (6.5) does not depend on $\mathbf{t}$ and is the criterion for A-optimality from (6.4).

The fractional factorial that is G-optimal for interactions minimizes

$$\max_W \mathrm{var}_F(\Gamma_W) = \max_W \{\mathrm{v}_W - \frac{\mathrm{v}_W^2}{\mathrm{v}_{(W\mathcal{A}_F)} + \sigma_p^2}\}. \tag{6.6}$$

6.5 c-Optimality

The $\mathbf{c}$-optimal fractional factorial minimizes the posterior variance of $\mathbf{c}'\mathbf{X}$ for some pre-specified vector $\mathbf{c}$. 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 $\mathbf{c}$ is the vector of 1's 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 the overall effect, $\Gamma_\emptyset$. So for $\mathbf{c} = \mathbf{1}$ the 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 $\sum_{W\in\mathcal{A}_F} \mathrm{v}_W$. This criterion can be restated in a more intuitive form. We have

$$\begin{aligned}\sum_{W\in\mathcal{A}_F} \mathrm{v}_W = \mathrm{var}(\Gamma_{\mathcal{A}_F}) &= E[\sum_{\mathbf{f}\in F} X(\mathbf{f}) \sum_{\mathbf{h}\in F} X(\mathbf{h})] = \sum_{\mathbf{f}\in F}\sum_{\mathbf{h}\in F} EX(\mathbf{f})X(\mathbf{h}) \\ &= \sum_{\mathbf{f}\in F}\sum_{\mathbf{h}\in F} EX(\mathbf{1})X(\mathbf{fh}) = 2^{k-p}\sum_{\mathbf{f}\in F} R(\mathbf{1},\mathbf{f}).\end{aligned} \tag{6.7}$$

Examining (6.7) shows that minimizing $\sum_{W\in\mathcal{A}_F} \mathrm{v}_W$ 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.

CHAPTER 7

Optimal Fractions for Orderly Processes

We propose orderly processes to be best-suited to most experimental situations. As defined in Chapter 3, orderly processes are stationary processes that have nested decreasing interaction variances and nested decreasing covariances. Informally, orderly processes have less dependence between runs that differ on more factors.

The results in this chapter are primarily meant to assure us that our approach produces reasonable designs. We first show in Section 7.1 that certain designs can always be excluded under D-optimality. In Section 7.2 we move on to the well-understood case of half-fractions. There is little debate about the best half-fraction design, and we show it is optimal in our framework.

We note that the only property of orderly processes needed for all of the results in this chapter is that they have nested decreasing interaction variances: $W \subset U \implies \mathrm{v}_W > \mathrm{v}_U$. None of the proofs requires the other property that defines orderly processes, namely that they have nested decreasing covariances.

7.1 D-Optimality

For orderly processes, certain designs can be immediately excluded from consideration for D-optimality. Lemma 7.1.1 limits the search for D-optimal designs to designs that use all k factors in the defining relations. These are generally considered the best designs in other frameworks. For example, the maximum resolution–minimum aberration design always uses all k factors in its defining relations (Fries and Hunter (1980)).

Lemma 7.1.1 *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 relation. Then there exists a fraction of the same size F^+ such that $|R_{F^+}+\sigma^2 I| > |R_F+\sigma^2 I|$ for $\sigma^2 \geq 0$. In particular, F is not D-Optimal.*

The proof of Lemma 7.1.1 is deferred to an appendix.

7.2 Half-Fractions

Half-fractions are well understood since they are defined by a single, non-empty 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 the Bayesian approach.

7.2.1 With or Without Error: D-, A-, and c-Optimality

We begin with a lemma that gives a partial ordering of half-fractions with respect to A-optimality. The proof of Lemma 7.2.1 is deferred to an appendix.

Lemma 7.2.1 *Suppose X is 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 $1 = W_l$ and F' be the half-fraction defined by $1 = W_{l-1}$. Then $\sum_{\mathbf{t}} var_F(X(\mathbf{t})) < \sum_{\mathbf{t}} var_{F'}(X(\mathbf{t}))$.*

Gathering this and previous results, we have the following proposition.

Proposition 7.2.1 *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 7.1.1 and A-optimality follows from Lemma 7.2.1 and Proposition 6.3.1. For the c-optimality result, recall from Section 6.5 that the best design for predicting $\sum_{\mathbf{t}} X(\mathbf{t})$ minimizes $\sum_{W \in \mathcal{A}_F} \mathrm{v}_W$. In the case of half-fractions, this means choosing the word W with the smallest interaction variance v_W. This is accomplished by choosing W to be the word of length k.□

7.2.2 Without Error: G- and E-Optimality

We can improve Proposition 7.2.1 when X is observed without error. Consider G-optimality for interactions.

Lemma 7.2.2 *Suppose X is 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 $\emptyset = W_{l-1}$. Consider any interaction Γ_W. Then there exists an interaction $\Gamma_{W'}$ such that $var_F(\Gamma_W) < var_{F'}(\Gamma_{W'})$.*

Proof: Say $W_l = LW_{l-1}$, i.e. W_{l-1} differs from W_l only by the factor L. According to (5.4), when $\sigma^2 = 0$ the posterior variance of an interaction Γ_W for X conditioned on a half-fraction is $\frac{\mathrm{v}_W \mathrm{v}_U}{\mathrm{v}_W + \mathrm{v}_U}$, 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 $L \in W_l W$.

If W does not contain L, then compare $\mathrm{var}_F(\Gamma_W)$ to $\mathrm{var}_{F'}(\Gamma_W) = \frac{\mathrm{v}_W \mathrm{v}_{W_{l-1}W}}{\mathrm{v}_W + \mathrm{v}_{W_{l-1}W}}$. 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, $\mathrm{v}_{W_l W} < \mathrm{v}_{W_{l-1}W} \implies \mathrm{var}_F(\Gamma_W) < \mathrm{var}_{F'}(\Gamma_W)$.

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

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 effect any subset or superset relationships.□

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

Lemma 7.2.3 *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 (6.6) and setting $\sigma^2 = 0$, $\mathrm{var}_F(\Gamma_W) = \mathrm{var}_F(\Gamma_U) = -\mathrm{cov}_F(\Gamma_W, \Gamma_U) = \frac{\mathrm{v}_W \mathrm{v}_U}{\mathrm{v}_W + \mathrm{v}_U}$ 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 $2\frac{\mathrm{v}_W \mathrm{v}_U}{\mathrm{v}_W + \mathrm{v}_U}$. 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 6.3.1 that E-optimality for runs and interactions are equivalent criteria. □

Bringing these results together, we have:

Proposition 7.2.2 *Suppose X is a stationary process with nested decreasing interaction variances observed without error. (For example, suppose X is an orderly process.) 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. Thus G-optimality for runs is a corollary of Proposition 7.2.1. G-optimality for interactions is a corollary of Lemma 7.2.2, which gives E-optimality by Lemma 7.2.3. □

CHAPTER 8

Asymptotically D-Optimal Fractions

There are an infinite number of distributions for a stationary process X on the k-dimensional space T and only finite number of fractional factorials of a given size. One can then 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. Realistically, one can hope 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 this chapter we consider only D-optimality. Results are derived for the most general family of stationary processes of interest, namely orderly processes. Because it is both instructive and an interesting 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. First, a distribution is manipulated in some way such that it converges to either complete independence (no correlation between $X(\mathbf{t})$ and $X(\mathbf{s})$ for $\mathbf{t} \neq \mathbf{s}$) or complete dependence (a degenerate process on T with perfect correlation from run to run). Taking limits, we are able to derive criteria for asymptotic D-optimality. The operations

considered in Sections 8.1 and 8.2 bring distributions toward complete run-to-run independence. Sections 8.3 and 8.4 bring distributions toward a degenerate process with complete dependence from run to run.

The operations used to manipulate distributions are exactly those considered in Chapter 4. There it was shown that a process that is stationary, isotropic, orderly, or tame remains so after manipulation. So these operations do not remove processes from the families to which they belong.

Note that we operate on the distribution of X, not $X + \epsilon$ when there is an error process. This keeps the magnitude of error separate and one is free to consider highly dependent processes and highly independent processes with both small and large error.

Without loss of generality, assume the covariance R for X on T is normalized to be a correlation matrix.

8.1 Toward Independence: Direct Powers

For a correlation R on T, consider $R^{\otimes n}$, the direct product of R with itself n times. This method of moving a correlation function toward independence was studied in Johnson, Moore, and Ylvisaker (1990). The D-optimal fractional factorial maximizes $|R_F^{\otimes n} + \sigma^2 I|$. Repeating the proof in Johnson, Moore, and Ylvisaker ((1990), (4.2)) to include the possibility of observational error ($\sigma^2 > 0$), the principal terms in the determinant of $R_F^{\otimes n} + \sigma^2 I$ 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 $\sum_{\mathbf{f} \neq \mathbf{f}' \in F} [R(\mathbf{f}, \mathbf{f}')]^{2n}$. (Notice the same results follows whether $\sigma^2 = 0$ or $\sigma^2 > 0$.) Thus the asymptotically 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.

8.2 Toward Independence: Convex Combinations

We can bring a correlation R toward independence in a different way, by setting $R^\lambda = \lambda R + (1-\lambda)I$ and letting $\lambda \to 0$. Including the case of error-free observation (set $\sigma^2 = 0$), the determinant of the correlation matrix for a design F is $|R_F^\lambda + \sigma^2 I| = |\lambda R_F + (1-\lambda)I + \sigma_p^2 I| = (1+\sigma_p^2)^{|F|} - \lambda^2 \sum_{\mathbf{f}\neq\mathbf{f}'\in F}[R(\mathbf{f},\mathbf{f}')]^2 + o(\lambda^2)$. 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

$$\sum_{i=1}^{k} D_i r_i^2. \tag{8.1}$$

8.3 Toward Dependence: Matrix Powers

Matrix powers R^n of a correlation matrix R must be normalized to retain "1" along the diagonal. We can use interaction variances to calculate a diagonal element of R^n: $R^n(\mathbf{t},\mathbf{t}) = 2^{-k}\text{trace}(R^n) = 2^{nk-k}\text{trace}(V^n) = 2^{nk-k}\sum_W \mathrm{v}_W^n$. Dividing every element of R^n by this factor, we know $\frac{2^{k-nk}}{\sum_W \mathrm{v}_W^n} R^n = 2^{2k} O'(\frac{1}{\sum_W \mathrm{v}_W^n} V^n) O$ is a correlation matrix.

Next, letting n get large, we show processes distributed as R^n tend to the degenerate process on T that is completely dependent from run to run. A diagonal element of $\frac{1}{\sum_W \mathrm{v}_W^n} V^n$ looks like $\frac{\mathrm{v}_U^n}{\sum_W \mathrm{v}_W^n}$. If the v_W are nested decreasing (in

particular, for an orderly process), then $\frac{\mathrm{v}_U^n}{\sum_W \mathrm{v}_W^n} = 1/\sum_W (\frac{\mathrm{v}_W}{\mathrm{v}_U})^n \to 0$ as $n \to \infty$ for all interaction variances *except* $\mathrm{v}_\emptyset$. So $\frac{2^{k-nk}}{\sum_W \mathrm{v}_W^n} R^n \to 2^{2k} O' V_J O$, where V_J is 0 every where 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.

Alternatively, 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 $\frac{2^{k-nk}}{\sum_W \mathrm{v}_W^n} R^n \to J$.

To omit unimportant constants and simplify presentation, write $R^{(n)}$ for the correlation $\frac{2^{k-nk}}{\sum_W \mathrm{v}_W^n} R^n$ and $\mathrm{v}_W^{(n)}$ for the interaction variances corresponding to $R^{(n)}$. 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 $W \supset U$. Similarly, $R_F^{(n)}$ means the principal submatrix of $R^{(n)}$ corresponding to the design F.

Now 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, so the two cases are presented separately.

8.3.1 Error-free Observation

First consider a process with correlation $R^{(n)}$ observed without error. The asymptotically D-optimal fractional factorial maximizes $\prod_{\mathcal{A}} \sum_{W \in \mathcal{A}} \mathrm{v}_W^{(n)}$, according to (6.3), for large n.

For a fractional factorial F, each alias set $\mathcal{A}$ has a largest interaction variance $\mathrm{v}_{\mathcal{A}^{\mathrm{M}}}$ occurring $n_{\mathcal{A}}$ times. For example, with an orderly process the alias set of defining relations $\mathcal{A}_F$ always has $\mathrm{v}_\emptyset$ as its unique largest interaction variance, and

$n_{\mathcal{A}_F} = 1$. Then as $n \to \infty$,

$$\prod_{\mathcal{A}} \sum_{W \in \mathcal{A}} \mathrm{v}_W^{(n)} \sim \prod_{\mathcal{A}} n_{\mathcal{A}} \mathrm{v}_{\mathcal{A}^{\mathrm{M}}}^{(n)}. \tag{8.2}$$

It is easiest to understand (8.2) with an example.

Example 5 Consider 2^{6-2} fractional factorials and an orderly isotropic processes. The maximum resolution fraction is F_1, defined by $\emptyset = 1234 = 3456 = 1256$, and

$$\begin{aligned} D(F_1) &= (\mathrm{v}_0^{(n)} + 3\mathrm{v}_4^{(n)})(4\mathrm{v}_3^{(n)})^2(2\mathrm{v}_2^{(n)} + 2\mathrm{v}_4^{(n)})^6(\mathrm{v}_1^{(n)} + 2\mathrm{v}_3^{(n)} + \mathrm{v}_5^{(n)})^6(3\mathrm{v}_2^{(n)} + \mathrm{v}_6^{(n)}) \\ &\sim \mathrm{v}_0^{(n)}(4\mathrm{v}_3^{(n)})^2(2\mathrm{v}_2^{(n)})^6(\mathrm{v}_1^{(n)})^6(3\mathrm{v}_2^{(n)}) = 1536(\mathrm{v}_0\mathrm{v}_1^6\mathrm{v}_2^7\mathrm{v}_3^2)^{(n)} \end{aligned}$$

Let F_2 be the fraction with defining relations $\emptyset = 123 = 3456 = 12456$. Looking at (8.2),

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

Since $1536(\mathrm{v}_0\mathrm{v}_1^6\mathrm{v}_2^7\mathrm{v}_3^2)^{(n)}/8(\mathrm{v}_0\mathrm{v}_1^6\mathrm{v}_2^9)^{(n)} = 192(\mathrm{v}_3/\mathrm{v}_2)^{2n} \to 0$, 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. For the design F_1, seven alias sets contain multiple two-factor interactions, whereas two alias sets have no two-factor interactions and 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. If two designs are the same in this respect, then the coefficient $\prod_{\mathcal{A}} n_{\mathcal{A}}$ becomes important. When $n_{\mathcal{A}}$ interaction variances v_W of the same size are in the same alias set, $n_{\mathcal{A}}$ is counted in the coefficient $\prod_{\mathcal{A}} n_{\mathcal{A}}$. This coefficient therefore counts the number of different estimable models that use only lowest-order effects from each alias set. Loosely speaking, the criterion to maximize $\prod_{\mathcal{A}} n_{\mathcal{A}} \mathrm{v}_{\mathcal{A}^{\mathrm{M}}}$ favors designs that allow one to estimate models with the most low-order effects and, subject to doing so, the highest number of such models. It is similar to the criterion of *estimation capacity* (Cheng, Steinberg, and Sun (1999)), which seeks to maximize the number of estimable models that include all main effects and j two-factor interactions.

8.3.2 Observation with Error

Next suppose the process with correlation $R^{(n)}$ 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 $\sigma_p^{2 \cdot 2^{k-p}} + \sigma_p^{2(2^{k-p}-1)} \mathrm{v}_\emptyset^{(n)} + \sigma_p^{2(2^{k-p}-2)} \mathrm{v}_\emptyset^{(n)} \sum_{W \notin \mathcal{A}_F} \mathrm{v}_W^{(n)}$. So for large n the D-optimal design should maximize $\sum_{W \notin \mathcal{A}_F} \mathrm{v}_W^{(n)}$ or, equivalently, minimize $\sum_{W \in \mathcal{A}_F} \mathrm{v}_W^{(n)}$. In other words, the asymptotically D-optimal fraction minimizes the interaction variance v_W among words $W \neq \emptyset$ in $\mathcal{A}_F$.

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 shortest word length in the defining relations alias set and, subject to that condition, minimizes the number of words of that length. In other words, the

maximum resolution–minimum aberration design is asymptotically D-optimal.

8.4 Toward Dependence: Convex Combinations

In this section the 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.

8.4.1 Error-free Observation

The D-optimal fractional factorial maximizes $|\lambda R_F + (1-\lambda)J|$. Since R_F and J are simultaneously diagonalizable, this is equivalent to maximizing $|\lambda V_F + (1-\lambda)V_J|$. Recall from Section 8.3 that V_J is 0 everywhere except for a 1 in the diagonal position corresponding to $\mathrm{v}_{\mathcal{A}_F}$. So $|\lambda V_F + (1-\lambda)V_J| = (\lambda \mathrm{v}_{\mathcal{A}_F} + 1-\lambda)\prod_{\mathcal{A}\neq\mathcal{A}_F} \lambda \mathrm{v}_{\mathcal{A}} = \lambda^{|F|-1}\left[(1-\lambda)\frac{|V_F|}{\mathrm{v}_{\mathcal{A}_F}} + \lambda|V_F|\right]$. 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, the asymptotic criterion of 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 $\mathrm{v}_{\mathcal{A}_F} \leq \mathrm{v}_{\mathcal{A}_{F'}}$, or

$$\sum_{W \in \mathcal{A}_F} \mathrm{v}_W \leq \sum_{W \in \mathcal{A}_{F'}} \mathrm{v}_W. \tag{8.3}$$

In other words, for *any* $\lambda \in (0,1]$, F remains D-optimal for R^λ if (8.3) holds.

Minimizing $\sum_{W \in \mathcal{A}_F} \mathrm{v}_W$ was the criterion resulting in Section 6.5 for minimizing the prediction error of $\sum_{\mathbf{t}} X(\mathbf{t})$. Section 6.5 showed that $\sum_{W \in \mathcal{A}_F} \mathrm{v}_W = 2^{k-p} \sum_{\mathbf{f} \in F} R(\mathbf{1}, \mathbf{f})$. In the isotropic case, minimizing $\sum_{W \in \mathcal{A}_F} \mathrm{v}_W$ is the same as minimizing

$$\sum_{i=0}^{k} D_i r_i. \tag{8.4}$$

Notice (8.4) is very similar to (8.1).

8.4.2 Observation with Error

Again, $R^\lambda = \lambda R + (1-\lambda)J$ and we seek to maximize $|\lambda R_F + (1-\lambda)J + \sigma^2 I|$ as $\lambda \to 0$. 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

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

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 $\sigma_p^2 \mathrm{v}_{\mathcal{A}_F} + \sigma_p^2 \sum_{\mathcal{A} \neq \mathcal{A}_F} \mathrm{v}_{\mathcal{A}} + \sum_{\mathcal{A}} \mathrm{v}_{\mathcal{A}} - \mathrm{v}_{\mathcal{A}_F} = (\sigma_p^2 + 1)\sum_{\mathcal{A}} \mathrm{v}_{\mathcal{A}} - \mathrm{v}_{\mathcal{A}_F}$. 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 6.5 and Section 8.4.1. In the isotropic case, the criterion can be restated as minimizing $\sum D_i r_i$, as at (8.4).

8.5 Discussion of Asymptotic Criteria

Table 8.1 summarizes the asymptotic criteria for D-optimality derived in sections 8.1 through 8.4. For simplicity, the criteria are presented for orderly isotropic processes.

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 in the same way. It is not surprising, then, that the approach toward independence brings about a maximin

criterion on runs in the design whereas the the approach toward dependence brings about a maximin criterion on word lengths (in the case $\sigma^2 > 0$).

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 $\prod_{\mathcal{A}} n_{\mathcal{A}}{}^{\mathrm{V}_{\mathcal{A}}{}^{\mathrm{M}}}$
$\sigma^2 > 0$	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 $\sum D_i r_i$

Table 8.1: D-Optimal Criteria for Orderly Isotropic Processes

Studying tables of designs for various k and p has led to an observation about the relationship between the maximin distance criterion, the maximin word length criterion, and minimizing the design correlation, $\sum D_i r_i$. For an example, refer to Tables 8.2 and 8.3, which give all 2^{8-3} fractional factorials of resolution III or higher. The designs are listed 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 list and compare the design correlation $\sum D_i r_i$. Subtracting the design correlation for the first design from the design correlation for the second design, one has $(2r_2 + 8r_3 + 10r_4 + 18r_5 + 2r_6 + r_8) - (r_2 + 10r_3 + 11r_4 + 4r_5 + 3r_6 + 2r_7) = D^4 r_2 + 2D^4 r_3 + D^4 r_4$. If the r_i are completely monotone, then $D^4 r_2, D^4 r_3$, and $D^4 r_4$ are all positive so this difference is always

positive. Therefore, for any tame process, the first design always has smaller design correlation.

This sort of analysis holds for other designs. For example, comparing the design correlation for the second and third designs in the list one has $(4r_2 + 22r_4 + 4r_6 + r_8) - (2r_2 + 8r_3 + 10r_4 + 18r_5 + 2r_6 + r_8) = 2D^4 r_2$, which is positive for completely monotone r_i. 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 fourth and fifth designs on the list. We have $(r_1 + 7r_3 + 14r_4 + 7r_5 + r_7 + r_8) - (2r_2 + 9r_3 + 9r_4 + 6r_5 + 4r_6 + r_7) = r_1 - 2r_2 - 2r_3 + 5r_4 + r_5 - 4r_6 + r_8$. To show that this quantity is not consistently positive or negative for completely monotone sequences of r_i, set $r_i = \rho^i$. For $\rho = 0.25$, this difference is positive (0.1133), but for $\rho = 0.75$, this difference is negative (-0112). Therefore, it cannot be written as the sum of difference operators D^i on the r_i.

Observations such as these leads to the following conjecture.

Conjecture 1 *Let F and F' be 2^{k-p} fractional factorials for some k and p, with respective design distances D_i and D'_i and respective word lengths L_i and L'_i. Suppose F is better than F' according to both the maximin word length criterion (resolution–aberration) and the maximin design distance criterion. Then 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 follow with an illustration of how these asymptotic criteria can be used to evaluate fractional factorials.

Example 6 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 Tables 8.2 and 8.3, which give all the 2^{8-3} fractional factorials of resolution III or higher. 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 (Proposition 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 $\prod_{\mathcal{A}} n_{\mathcal{A}} \mathrm{v}_{\mathcal{A}}{}^{\mathrm{M}}$ for any orderly isotropic prior distribution. For example, comparing the first and second designs, $2^9 3^2 4^4 5^4 (\mathrm{v}_1^8 \mathrm{v}_2^{15} \mathrm{v}_3^8)^n / 2^6 3^3 4^1 (\mathrm{v}_1^8 \mathrm{v}_2^{20} \mathrm{v}_3^3)^n = 2^9 5^4 \mathrm{v}_3^{5n} / 3 \mathrm{v}_2^{5n}$ goes to 0 as $n \to \infty$, so the first design is better.

Altogether, for every criterion in Table 8.1 (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 distributions both nearly-dependent and nearly-independent on T, the first design is optimal. □

Example 7 Chen (1998) gives the maximum resolution–minimum aberration 2^{14-7} design (i.e. the maximin word length design), which is generated by $\emptyset = 1238 = 4569 = 1245\overline{10} = 1346\overline{11} = 1246\overline{12} = 2345\overline{14} = 1356\overline{13}$. It turns out that the vector of design distances is the same as the vector of word lengths for this design (which is possible since the design and its dual are the same size). Therefore this is also the maximin distance design. In addition, if Conjecture 1

is correct then this design also minimizes $\sum D_i r_i$ and $\sum D_i r_i^2$, assuming r_i is completely monotone, among all 2^{14-7} fractional factorials.□

Example 8 Table 8.4 gives two 2^{11-5} fractional factorials. The first design in Table 8.4 is the maximum resolution–minimum aberration (maximin word length) fraction (Franklin (1980)). However, we see that it is *not* the maximin distance design. The second design in Table 8.4 has minimum distance 4, as opposed to 3 for the first design. However, $\frac{2^{15}5^{12}}{2^{10}3^{1}4^{6}6^{1}}(\frac{\mathrm{v}_1^{11}\mathrm{v}_2^{40}\mathrm{v}_3^{12}}{\mathrm{v}_1^{11}\mathrm{v}_2^{44}\mathrm{v}_3^{8}})^n \sim \frac{5^{12}}{2^{8}3^{2}}(\frac{\mathrm{v}_3}{\mathrm{v}_2})^{4n} \to 0$, so the maximum resolution–minimum aberration design beats the second design in maximizing $\prod_{\mathcal{A}} n_{\mathcal{A}^{\mathrm{V}}\mathcal{A}^{\mathrm{M}}}$.

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 8.1 graphs this difference for the family of correlations $r_i = \rho^i$, with $0 \leq \rho \leq 1$. 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 correspond with 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 on this criterion.

In this example, the first design is most appropriate for experiments modeled with a process with high run-to-run dependence because it is maximin word length, maximizes $\prod_{\mathcal{A}} n_{\mathcal{A}^{\mathrm{V}}\mathcal{A}^{\mathrm{M}}}$, 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 is maximin distance and minimizes $\sum D_i \rho^i$ for ρ near 0. □

In looking at designs for different values of k and p, we have seen one of two scenarios in each case. Either (1) the maximum resolution–minimum aberra-

tion design always turns out optimal, or (2) the maximum resolution–minimum aberration design is optimal for some asymptotic criteria, but another design is optimal for other asymptotic criteria. In our experience, the most common criterion for which the maximum resolution–minimum aberration design is sub-optimal is the criterion to maximize $\prod_{\mathcal{A}} n_{\mathcal{A}^{\mathrm{V}}\mathcal{A}^{\mathrm{M}}}$. The designs that are best for this criterion are the same designs that have been proposed by other authors to be better than the maximum resolution–minimum aberration design. For example, Chen, Sun, and Wu (1993) note that the 2^{6-2} design called F_2 in Example 6 might be better for some experiments than the maximum resolution–minimum aberration design because it allows independent estimates of some two-factor interactions. Altogether, the designs that appear in our framework are reasonable, and are designs that have been proposed under other criteria.

Defining Relations	Word Lengths							
	Design Distances							
	$\prod_{\mathcal{A}} n_{\mathcal{A}^{\mathrm{V}}\mathcal{A}^{\mathrm{M}}}$							
	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	$2^{12} 4^9 8^2 \mathrm{V}_1^8 \mathrm{V}_2^{13} \mathrm{V}_3^8 \mathrm{V}_4^2$							
1234	0	0	0	7	0	0	0	0
1256	1	0	7	14	7	0	1	1
1357	$3^{14} 7^2 \mathrm{V}_1^8 \mathrm{V}_2^{14} \mathrm{V}_3^8 \mathrm{V}_4$							
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	$2^7 \mathrm{V}_1^8 \mathrm{V}_2^{19} \mathrm{V}_3^4$							

Table 8.2: 2^{8-3} Fractional Factorials, Part I

Defining Relations	Word Lengths / Design Distances / $\prod_{\mathcal{A}} n_{\mathcal{A}^{\mathrm{V}}\mathcal{A}^{\mathrm{M}}}$							
	1	2	3	4	5	6	7	8
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	2	3	2	0	0	0
145	1	1	7	11	7	3	1	0
12467	$2^{10} 3^2 \mathrm{V}_1^8 \mathrm{V}_2^{15} \mathrm{V}_3^8$							
123	0	0	3	1	0	2	1	0
456	0	5	4	7	12	3	0	0
1245678	$2^8 \mathrm{V}_1^8 \mathrm{V}_2^{17} \mathrm{V}_3^6$							
123	0	0	3	2	1	1	0	0
145	1	2	6	9	9	4	0	0
124567	$2^8 \mathrm{V}_1^8 \mathrm{V}_2^{15} \mathrm{V}_3^8$							
123	0	0	3	3	0	0	1	0
145	1	3	3	11	11	1	1	0
1234567	$2^{12} 4^4 \mathrm{V}_1^8 \mathrm{V}_2^{13} \mathrm{V}_3^8 \mathrm{V}_4^2$							
123	0	0	4	3	0	0	0	0
145	2	1	4	11	10	3	0	0
1246	$3^4 \mathrm{V}_1^8 \mathrm{V}_2^{14} \mathrm{V}_3^8 \mathrm{V}_4$							

Table 8.3: 2^{8-3} Fractional Factorials, Part II

Defining Relations	Word Lengths / Design Distances / $\prod_{\mathcal{A}} n_{\mathcal{A}}{}^{\mathrm{V}} \mathcal{A}^{\mathrm{M}}$										
	1	2	3	4	5	6	7	8	9	10	11
3457	0	0	0	4	14	8	0	3	2	0	0
2458	0	0	2	14	22	8	6	9	2	0	0
123469	$2^{10}3^{1}4^{6}6^{1}\mathrm{V}_{1}^{11}\mathrm{V}_{2}^{44}\mathrm{V}_{3}^{8}$										
$123\overline{510}$											
$145\overline{611}$											
34567	0	0	0	5	10	10	5	0	0	0	1
14568	0	0	0	25	0	27	0	10	0	1	0
12569	$2^{15}5^{12}\mathrm{V}_{1}^{11}\mathrm{V}_{2}^{40}\mathrm{V}_{3}^{12}$										
$123\overline{610}$											
$234\overline{611}$											

Table 8.4: 2^{11-5} Fractional Factorials

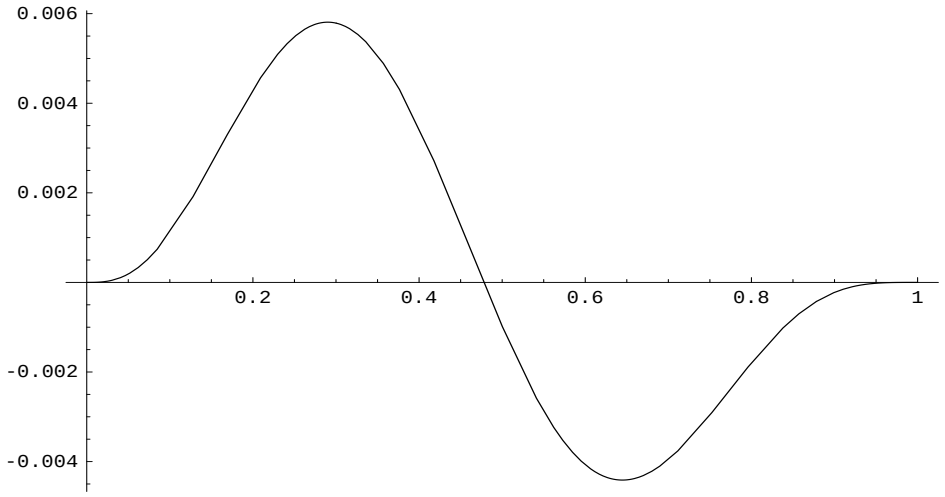

Figure 8.1: Difference of Design Correlations $\sum D_i \rho^i$ for Two 2^{11-5} Fractional Factorials, as a Function of ρ

APPENDIX A

Proof of Proposition 3.3.1

Proof: As in Section 3.2, let $\mathbf{r}$ denote the vector of covariances $(r_0, r_1, \ldots, r_k)'$ and $\mathbf{v}$ denote the vector of interaction variances $(\mathrm{v}_0, \mathrm{v}_1, \ldots, \mathrm{v}_k)'$ so an isotropic process on T. Proposition 3.3.1 states that if either of $\mathbf{v}$ or $\mathbf{r}$ is completely monotone, so is the other. Since the same transformation (aside from scale) takes $\mathbf{r}$ to $\mathbf{v}$ as takes $\mathbf{v}$ to $\mathbf{r}$, it suffices to prove one direction. Assume the v_j are completely monotone and prove the r_i are as well.

First, establish an identity for Krawtchouk polynomials. For $i = 0, 1, \ldots, k-1$ and $j = 1, \ldots, k$,

$$P_j(i;k) - P_j(i+1;k) = 2P_{j-1}(i;k-1). \tag{A.1}$$

The identity (A.1) is most easily derived using the generating functions for Krawtchouk polynomials: $\sum_{j=0}^{k} P_j(i;k)z^j = (1+z)^{k-i}(1-z)^i$. Call $g_{i,k}(z) = (1+z)^{k-i}(1-z)^i$.

$$\begin{aligned}
P_j(i;k) - P_j(i+1;k) &= g_{i,k}^{(j)}(0)/j! - g_{i+1,k}^{(j)}(0)/j! \\
&= [(1+z)^{k-i}(1-z)^i - (1+z)^{k-i-1}(1-z)^{i+1}]_{z=0}^{(j)}/j! \\
&= [(1+z)^{k-i-1}(1-z)^i(1+z-(1-z))]_{z=0}^{(j)}/j! \\
&= [2zg_{i,k-1}(z)]_{z=0}^{(j)}/j! \\
&= 2[jg_{i,k-1}^{(j-1)}(z) + zg_{i,k-1}^{(j)}(z)]_{z=0}/j! \\
&= 2g_{i,k-1}^{(j-1)}(0)/(j-1)! = 2P_{j-1}(i;k-1)
\end{aligned}$$

Consider first differences $D^1 r_i = r_i - r_{i+1}$, $i = 0, 1, \ldots, k-1$. The identity (A.1) implies $D^1 r_i = \sum_{j=0}^{k} [P_j(i;k) - P_j(i+1;k)] \mathrm{v}_j = 2 \sum_{j=1}^{k} P_{j-1}(i;k-1) \mathrm{v}_j$. Now,

$$\begin{aligned}
\sum_{j=1}^{k} P_{j-1}(i;k-1)\mathrm{v}_j &= \sum_{j=1}^{k} \sum_{m} (-1)^m \binom{i}{m} \binom{k-1-i}{j-1-m} \mathrm{v}_j \\
&= \sum_{m=0}^{i} (-1)^m \binom{i}{m} \sum_{m+1 \le j \le k+m-i} \binom{k-1-i}{j-1-m} \mathrm{v}_j \\
&= \sum_{m=0}^{i} (-1)^m \binom{i}{m} \sum_{1 \le q \le k-i} \binom{k-1-i}{q-1} \mathrm{v}_{q+m} \\
&= \sum_{q=1}^{k-i} \binom{k-1-i}{q-1} \sum_{m=0}^{i} (-1)^m \binom{i}{m} \mathrm{v}_{q+m} \\
&= \sum_{q=1}^{k-i} \binom{k-1-i}{q-1} D^i \mathrm{v}_q . \qquad \text{(A.2)}
\end{aligned}$$

Complete monotoneity of $\mathbf{v}$ means $D^i \mathrm{v}_q \ge 0$, so the final expression is non-negative.

A bit of notation helps finish the proof. Write $\mathbf{r_{D^m}}$ for the $k+1-m$ vector of $D^m r_i, i = 0, \ldots, k-m$ and $\mathbf{v_{(m)}}$ for the vector of the last $k+1-m$ components of $\mathbf{v}$ (so $\mathbf{r_{D^0}} = \mathbf{r}$ and $\mathbf{v_{(0)}} = \mathbf{v}$). The Krawtchouk matrix K_k transforms $\mathbf{v_{(0)}}$ to $\mathbf{r_{D^0}}$, $\mathbf{r_{D^0}} = K_k \mathbf{v_{(0)}}$, and the proof so far has shown $\mathbf{r_{D^1}} = K_{k-1} \mathbf{v_{(1)}}$ Note that $\mathbf{r_{D^{m+1}}}$ is gotten from $\mathbf{r_{D^m}}$ in the same way that $\mathbf{r_{D^1}}$ came from $\mathbf{r_{D^0}}$: by taking differences in adjacent components. Inductively applying the identity (A.1) gives the relation $\mathbf{r_{D^m}} = 2^m K_{k-m} \mathbf{v_{(m)}}$. The last step is the calculation as at (A.2) that shows the inner product of the ith row of K_{k-m} with $\mathbf{v_{(m)}}$ is non-negative: $\sum_{j=m}^{k} P_{j-m}(i;k-m) \mathrm{v}_j = \sum_{q=m}^{k-i} \binom{k-m-i}{q-m} D^i \mathrm{v}_q \ge 0$.

APPENDIX B

Proof of Proposition 4.2.1

In this appendix, subscripts on processes as in X_1 and X_2 are used to differentiate processes, and not to suggest replications of a process observed with error as at (2.2).

Proposition 4.2.1 states that if X_1 and X_2 are orderly processes distributed as R_1 and R_2, respectively, then so is the process X distributed as $R_1 \otimes R_2$. Although this proposition arises in Chapter 4, it is convenient to use some notation and results for fractional factorials from later chapters.

The process X has covariance $R_1 \otimes R_2$, and Section 4.1 showed that $R_1 \otimes R_2$ is a legitimate covariance matrix for a stationary process on T. Clearly, taking direct products of the matrices R_1 and R_2 preserves the nested decreasing property on covariances. Also, since $R_1 \otimes R_2$ is a covariance matrix its eigenvalues must be non-negative. Since the interaction variances for X are proportional to the eigenvalues of $R_1 \otimes R_2$, they are positive and well-defined. The only remaining question is whether X has nested decreasing interaction variances.

First note that a process has nested decreasing interaction variances if and only if for all factors L, for any word W not containing L, $\mathrm{v}_W > \mathrm{v}_{WL}$. We assume X_1 and X_2 (R_1 and R_2) have this property and prove that X does as well.

Now for any factor L, consider separating the runs in T into two groups depending on whether t_L is 1 or -1. In fact, the first group is just a half-fraction

defined by $1 = L$. As a fractional factorial, it is a subgroup of T, and the set of runs for which $t_L = -1$ is a coset. Let $\mathbf{t}_L$ denote the run that is 1 in every position except for factor L, and denote the the two groups by F and $\mathbf{t}_L F$.

For $i = 1, 2$, order the vector of observations $X_i(T)$ as $[X_i(F)\,;\, X_i(\mathbf{t}_L F)]$, where the runs in $\mathbf{t}_L F$ are written in corresponding order to the runs in F. Notice that stationarity of X_i implies $R_{i,F} = \mathrm{var}(X_i(F)) = \mathrm{var}(X_i(\mathbf{t}_L F))$. Therefore we can write R_i in 2^{k-1} by 2^{k-1} blocks:

$$R_i = \begin{bmatrix} R_{i,F} & G_i \\ G_i' & R_{i,F} \end{bmatrix},$$

where $G_i = \mathrm{cov}(X_i(F), X_i(\mathbf{t}_L F))$.

Next, divide the set of interactions into two groups, interactions for words that do not contain L, $\Gamma_{\overline{L}}$. and interactions for words that contain L, Γ_L. Write the vector of interactions Γ as $[\Gamma_{\overline{L}}\,;\Gamma_L]$, and again assume that interactions in Γ_L are written in order corresponding to $\Gamma_{\overline{L}}$.

With these orderings, notice the matrix $O = O_k$ that transforms observations to interactions can also be written in 2^{k-1} by 2^{k-1} blocks,

$$O = \begin{bmatrix} O_{k-1} & O_{k-1} \\ O_{k-1} & -O_{k-1} \end{bmatrix}.$$

Moreover, the matrix O_{k-1} is the orthogonal matrix for $k-1$ factors, and is also the matrix O_F for the half-fraction F defined by $1 = L$.

The diagonal matrix V_i of interaction variances is OR_iO'. Using the partitions of O and R_i above, we have

$$V_i = \begin{bmatrix} O_F(R_{i,F} + G_i)O_F' & 0 \\ 0 & O_F(R_{i,F} - G_i)O_F' \end{bmatrix}. \tag{B.1}$$

Since V_i and $O_F R_{i,F} O_F'$ are both diagonal, $O_F G_i O_F'$ must also be diagonal.

The claim is that the matrix G_i can serve as a covariance matrix for a stationary process $\overline{X}$ for 2^{k-1} factors. To prove this claim we must prove that G_i is symmetric, stationary, and positive definite.

Since runs $\mathbf{f}$ in F all have $t_L = 1$, there is no difficulty in thinking of them as runs in 2^{k-1} by ignoring the factor L component, in which case we denote them by $\overline{\mathbf{f}}$. A generic entry in G_i is $\mathrm{cov}(X_i(\mathbf{t}), X_i(\mathbf{t}_L\mathbf{s}))$, where $\mathbf{t}$ and $\mathbf{s}$ are in F. As a covariance matrix for $\overline{X}_i$ on 2^{k-1}, let this entry represent $\mathrm{cov}(\overline{X}_i(\overline{\mathbf{t}}), \overline{X}_i(\overline{\mathbf{s}}))$.

Let $\mathbf{s}$ and $\mathbf{t}$ be runs in F and say $\mathbf{ts} = \mathbf{u} \in F$. The cross-diagonal entry in G_i of $\mathrm{cov}(X_i(\mathbf{t}), X_i(\mathbf{t}_L\mathbf{s}))$ is $\mathrm{cov}(X_i(\mathbf{s}), X_i(\mathbf{t}_L\mathbf{t}))$. Because X_i is stationary, $\mathrm{cov}(X_i(\mathbf{s}), X_i(\mathbf{t}_L\mathbf{t})) = \mathrm{cov}(X_i(\mathbf{su}), X_i(\mathbf{t}_L\mathbf{tu})) = \mathrm{cov}(X_i(\mathbf{t}), X_i(\mathbf{t}_L\mathbf{s}))$, so G is symmetric. Also notice that for $\mathbf{f} \in F$, $\mathrm{cov}(\overline{X}_i(\overline{\mathbf{t}}), \overline{X}_i(\overline{\mathbf{s}})) = \mathrm{cov}(X_i(\mathbf{t}), X_i(\mathbf{t}_L\mathbf{s})) = \mathrm{cov}(X_i(\mathbf{tf}), X_i(\mathbf{t}_L\mathbf{sf})) = \mathrm{cov}(\overline{X}_i(\overline{\mathbf{tf}}), \overline{X}_i(\overline{\mathbf{sf}}))$, so G_i is stationary.

Now, G_i is positive definite if and only if its eigenvalues are positive. The eigenvalues of G_i are proportional to the diagonal entries in $O_F G_i O_F'$. Examine (B.1) and recall that every diagonal entry in $O_F(R_{i,F} + G_i)O_F'$ represents v_W for some W not containing L, whereas the corresponding entry in $O_F(R_{i,F} - G_i)O_F'$ represents v_{WL}. The assumption that X_i has nested decreasing interaction variances implies that for all diagonal entries, $O_F(R_{i,F} + G_i)O_F' > O_F(R_{i,F} - G_i)O_F'$. Therefore the diagonal entries of $O_F G_i O_F'$ must be positive.

So we can let $\overline{X}_1$ and $\overline{X}_2$ be independent stationary process on 2^{k-1} with respective covariances G_1, and G_2. Then there exists a stationary process $\overline{X}$ on 2^{k-1} distributed as $G_1 \otimes G_2$. In particular, $G_1 \otimes G_2$ has positive eigenvalues and is diagonalized by O_F.

Finally, the covariance $R = R_1 \otimes R_2$ for X on 2^k has interaction variances

given by the diagonal matrix $O_F R O_F' =$

$$\begin{bmatrix} O_F(R_{1,F} \otimes R_{2,F} + G_1 \otimes G_2)O_F' & 0 \\ 0 & O_F(R_{1,F} \otimes R_{2,F} - G_1 \otimes G_2)O_F' \end{bmatrix}.$$

For the interaction variances to be nested decreasing, we must have $O_F(R_{1,F} \otimes R_{2,F} + G_1 \otimes G_2)O_F' > O_F(R_{1,F} \otimes R_{2,F} - G_1 \otimes G_2)O_F'$ for diagonal entries. This is true if and only if $O_F(G_1 \otimes G_2)O_F'$ has positive diagonal entries, which is the case since $G_1 \otimes G_2$ is a positive-definite matrix and diagonalized by O_F.

APPENDIX C

Proof of Proposition 4.4.1

Let r_i and s_i, $i = 0, \ldots, k$ be completely monotone. Proposition 4.4.1 states that $r_i s_i$ is again completely monotone.

Choose m, j such that $m + j \leq k$. Let $0 \leq l \leq m$. By completely monotoneity of the two sequences r_i and s_i,

$$
\begin{aligned}
0 \leq & \left[\sum_{p=0}^{l} \binom{l}{p} D^{m-l} r_{j+p}\right] \left[\sum_{p=0}^{m-l} \binom{m-l}{p} D^{l} s_{j+p}\right] \\
= & \left[\sum_{p=0}^{l}\sum_{h=0}^{m-l} (-1)^h \binom{l}{p}\binom{m-l}{h} r_{j+p+h}\right] \left[\sum_{p=0}^{m-l}\sum_{h=0}^{l} (-1)^h \binom{m-l}{p}\binom{l}{h} s_{j+p+h}\right] \\
= & \left[\sum_{x=0}^{m} r_{j+x} \sum_{q=0}^{x} (-1)^{x-q} \binom{l}{q}\binom{m-l}{x-q}\right] \left[\sum_{y=o}^{m} s_{j+y} \sum_{q=0}^{y} (-1)^{q} \binom{l}{q}\binom{m-l}{y-q}\right] \\
= & \left[\sum_{x=0}^{m} r_{j+x} (-1)^x P_x(l;m)\right] \left[\sum_{y=0}^{m} s_{j+x} P_y(l;m)\right]. \qquad \text{(C.1)}
\end{aligned}
$$

We use the non-negativity of (C.1) and the orthogonality of Krawtchouk polynomials with respect to binomial measure to show the m^{th} difference D^m of the

direct product sequence $r_i s_i$ is non-negative:

$$
\begin{aligned}
0 &\le \sum_{l=0}^{m} \binom{m}{l} \left[\sum_{x=0}^{m} r_{j+x} (-1)^x P_x(l;m) \right] \left[\sum_{y=0}^{m} s_{j+x} P_y(l;m) \right] \\
&= \sum_{x=0}^{m} \sum_{y=0}^{m} r_{j+x} s_{j+y} (-1)^x \left[\sum_{l=0}^{m} \binom{m}{l} P_x(l;m) P_y(l;m) \right] \\
&= \sum_{x=0}^{m} \sum_{y=0}^{m} r_{j+x} s_{j+y} (-1)^x \left[2^m \binom{m}{x} \delta_{xy} \right] = 2^m \sum_{x=0}^{m} r_{j+x} s_{j+x} (-1)^x \binom{m}{x} \\
&= 2^m D^m(r_j s_j).
\end{aligned}
$$

APPENDIX D

Proof of Proposition 5.1.1

For a subgroup F of T with alias set $\mathcal{A}$, the alias interaction $\Gamma_{\mathcal{A}}$ is defined as $\sum_{W\in\mathcal{A}}\Gamma_W$. Proposition 5.1.1 states that $2^k\Gamma_{\mathcal{A}} = \sum_{\mathbf{t}\in F}\sum_{W\in\mathcal{A}} X(\mathbf{t})W(\mathbf{t})$. Before proving the proposition, we state and prove a lemma.

Lemma D.0.1 *Let F be a subgroup of T, $\mathbf{t}_0$ be any run not in F, and $\mathcal{A}$ be an alias set for F. Then $\sum_{W\in\mathcal{A}} W(\mathbf{t}_0) = 0$.*

Proof: First consider $\mathcal{A}_F$. Since $\mathbf{t}_0 \notin F$ there is a word W^- in $\mathcal{A}_F$ such that $W^-(\mathbf{t}_0) = -1$. But $\emptyset \in \mathcal{A}_F$ and $\emptyset(\mathbf{t}_0) = 1$. Define $\mathcal{A}_F^{\mathbf{t}_0+}$ to be all the words W in $\mathcal{A}_F$ such that $W(\mathbf{t}_0) = 1$ and $\mathcal{A}_F^{\mathbf{t}_0-}$ to be all W in $\mathcal{A}_F$ such that $W(\mathbf{t}_0) = -1$. Neither $\mathcal{A}_F^{\mathbf{t}_0+}$ nor $\mathcal{A}_F^{\mathbf{t}_0-}$ is empty. Define two maps, $\mathcal{A}_F^{\mathbf{t}_0+} \to \mathcal{A}_F^{\mathbf{t}_0-} : W \to W\cdot W^-$ and $\mathcal{A}_F^{\mathbf{t}_0-} \to \mathcal{A}_F^{\mathbf{t}_0+} : W \to W\cdot W^-$. Since both maps are injective, $|\mathcal{A}_F^{\mathbf{t}_0+}| = |\mathcal{A}_F^{\mathbf{t}_0-}|$. This proves the lemma for $\mathcal{A}_F$.

For any other alias set $\mathcal{A}$, take $W \in \mathcal{A}$ and write $\mathcal{A} = W\mathcal{A}_F$. If $W(\mathbf{t}_0) = 1$ then $\mathcal{A}^{\mathbf{t}_0+} = \mathcal{A}_F^{\mathbf{t}_0+}W$ and $\mathcal{A}^{\mathbf{t}_0-} = \mathcal{A}_F^{\mathbf{t}_0-}W$. If $A(\mathbf{t}_0) = -1$ then $\mathcal{A}^{\mathbf{t}_0+} = \mathcal{A}_F^{\mathbf{t}_0-}A$ and $\mathcal{A}^{\mathbf{t}_0-} = \mathcal{A}_F^{\mathbf{t}_0+}A$. In either case $|\mathcal{A}^{\mathbf{t}_0+}| = |\mathcal{A}^{\mathbf{t}_0-}|$ and the lemma holds. □

The proof of Proposition 5.1.1 follows easily:

$$2^k \Gamma_{\mathcal{A}} = 2^k \sum_{W \in \mathcal{A}} \Gamma_W = \sum_{W \in \mathcal{A}} \sum_{\mathbf{t} \in T} W(\mathbf{t}) X(\mathbf{t}) = \sum_{\mathbf{t} \in T} X(\mathbf{t}) \sum_{W \in \mathcal{A}} W(\mathbf{t})$$
$$= \sum_{\mathbf{t} \in F} X(\mathbf{t}) \sum_{W \in \mathcal{A}} W(\mathbf{t}) + \sum_{\mathbf{t} \notin F} X(\mathbf{t}) \sum_{W \in \mathcal{A}} W(\mathbf{t}) \qquad \text{(D.1)}$$

The second term of (D.1) vanishes by Lemma D.0.1.

APPENDIX E

Proof of Lemma 7.1.1

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*

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

Proof:

$$(a+d)(b+c) \geq (a+c)(b+d) \iff ac+bd \geq ad+bd \iff a(c-d) \geq b(c-d)$$

$$(a+c)(b+d) \geq (a+b)(c+d) \iff ab+cd \geq ac+bd \iff a(b-c) \geq d(b-c)$$

□

Given four non-negative numbers $a \geq b \geq c \geq d$, in the context of pairing the numbers and producing the product of sums of pairs, 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 relation. Lemma 7.1.1 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 $W_1, W_2, ..., W_p$ 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 $W_1, W_2, ..., W_{p-1}, W_pL$. Referring to the D-optimality criterion at (6.3), 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 of 2^p words, $< W_1, \ldots, W_p >= \mathcal{A}_F$, is all the defining relations for F. The defining relations for F^+ is 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 the words in S in the natural order according to the given listing of its generators:

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

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 of the proof 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 $WL \in S_4$. $L \notin W \implies W \subset WL \implies \mathrm{v}_W > \mathrm{v}_{WL}$. Similarly, every word $W \in S_2$ does not contain L so $W \subset WL \in S_3 \implies \mathrm{v}_W > \mathrm{v}_{WL}$. 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 $UW \subset UWL \implies \mathrm{v}_{UW} > \mathrm{v}_{UWL}$, and so on.

If $L \in U$ only a minor adjustment to the argument is needed. For every word $UW \in S_1$, $UWL \in S_4$, $UW \supset UWL$. For every word $UW \in S_2$ $UWL \in S_3$, $UW \supset UWL$. 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 $\mathrm{v}'_W = \mathrm{v}_W + \frac{\sigma_p^2}{2^p}$. Since the v_W are nested decreasing, the v'_W are nested decreasing. Since D-optimality means maximizing $\prod_{\mathcal{A}}(\sigma_p^2 + \sum_{W \in \mathcal{A}} \mathrm{v}_W)$ and this equals $\prod_{\mathcal{A}} \sum_{W \in \mathcal{A}} \mathrm{v}'_W$, the same proof holds using v'_W in place of v_W.

APPENDIX F

Proof of Lemma 7.2.1

The proof of Lemma 7.2.1 requires another observation, which we leave to the reader to show.

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 $\frac{a^2+d^2}{a+d+s} + \frac{b^2+c^2}{b+c+s}$ as *best*, $\frac{a^2+c^2}{a+c+s} + \frac{b^2+d^2}{b+d+s}$ as *middle*, $\frac{a^2+b^2}{a+b+s} + \frac{c^2+d^2}{c+d+s}$ 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 7.2.1 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 (6.4), the proof shows $A(F) > A(F')$.

Partition the group of words into cosets of the subgroup $S =< W_{l-1}, W_l >= \{\emptyset, W_{l-1}, W_l, L\}$. 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 $\frac{\mathrm{v}_U^2+\mathrm{v}_{UW_l}^2}{\mathrm{v}_U+\mathrm{v}_{UW_l}+\sigma_p^2}+\frac{\mathrm{v}_{UW_{l-1}}^2+\mathrm{v}_{UL}^2}{\mathrm{v}_{UW_{l-1}}+\mathrm{v}_{UL}+\sigma_p^2}$. The corresponding terms for $A(F')$ are $\frac{\mathrm{v}_U^2+\mathrm{v}_{UW_{l-1}}^2}{\mathrm{v}_U+\mathrm{v}_{UW_{l-1}}+\sigma_p^2}+\frac{\mathrm{v}_{UW_l}^2+\mathrm{v}_{UL}^2}{\mathrm{v}_{UW_l}+\mathrm{v}_{UL}+\sigma_p^2}$.

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

[1] Chaloner, K. (1984). Optimal Bayesian Experimental Design for Linear Models. The Annals of Statistics **12**, 283-300.

[2] Chen, J. (1998). Intelligent Search for 2^{13-6} and 2^{14-7} Minimum Aberration Designs. Statistica Sinica **8**, 1265-1270.

[3] 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. International Statistical Review **61**, 131-145.

[4] Chen, J., and Wu, C.F.J. (1991). Some Results on s^{n-k} Fractional Factorial Designs with minimum Aberration or Optimal Moments. The Annals of Statistics **19**, 2, 1028-1041.

[5] Cheng, C.S., Steinberg, D.M., and Sun, D.X. (1999). Minimum Aberratin and Model Robustness for Two-Level Fractional Factorial Designs. Journal of the Royal Statistical Society, Series B **61**, 85-93.

[6] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments. Journal of the American Statistical Association **86**, 953-963.

[7] 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.

[8] Franklin, M.F. (1985). Selecting Defining Contrasts and Confounded Effects in p^{n-m} Factorial Experiments. Technometrics **27**, 2, 165-172.

[9] Fries, A. and Hunter, W.G. (1980). Minimum Aberration 2^{n-k} Designs. Technometrics **26**, 601-608.

[10] Johnson, M.E., Moore, L.M., and Ylvisaker, D. (1990). Minimax and Maximin Distance Designs. Journal of Statistical Planning and Inference **26**, 131-148.

[11] Laycock, P.J. and Rowley P.J. (1995). A Method for Generating and Labelling All Regular Fractions or Blocks for q^{n-m}-designs. Journal of the Royal Statistical Society **57**, 1, 191-204.

[12] MacWilliams, F.J. and Sloane, N.J.A. (1977). Theory of Error-Correcting Codes, North-Holland Publishing Company.

[13] Meyer, R.D., Steinberg, D.M., and Box, G. (1996). Follow-up Designs to Resolve Confounding in Multifactor Experiments. Technometrics **38**, 303-313.

[14] Mitchell, T.J., Morris, M.D., and Ylvisaker, D. (1995). Two-Level Fractional Factorials and Bayesian Prediction. Statistica Sinica **5**, 559-573.

[15] Mitchell, T., Sacks, J., and Ylvisaker, D. (1994). Asymptotic Bayes Criteria for Nonparametric Response Surface Design. The Annals of Statistics **22**, 634-651.

[16] Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H.P. (1989). Design and Analysis of Computer Experiments. Statistical Science **4**, 409-435.

[17] Tang, B. and Wu, C.F.J. (1996). Characterization of Minimum Aberration 2^{n-k} Designs in Terms of Their Complementary Designs. The Annals of Statistics **24**, 6, 2549-2559.

[18] Toman, B. (1994). Bayes Optimal Designs for Two- and Three-Level Factorial Experiments. Journal of the American Statistical Association **89**, 937-946.

[19] Widder, D.V. (1971). An Introduction to Transform Theory, Academic Press.