# Bias and mean square error in experimental designs

by

Leon Jordan Filho

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1971

TABLE OF CONTENTS

# I. INTRODUCTION

## A. GENERAL AIMS

This thesis is directed towards the general problem of design for estimating the relationship of a response variable to experimentally controlled variables. Chapters II, III, and IV discuss the case in which the control variable is continuous. Chapters V and VI deal with the case in which there are several qualitative control variables, which is the situation for which factorial experimentation was developed.

The set of prior possible values for the experimentally controlled variables is generally called factor space. The problem of design is to make a choice of points in the factor space at which to take observations, and a choice of how many observations to take at the chosen points. This problem is very difficult in general because almost any investigation has many aims and usually the aims conflict. The investigator will be concerned usually with making a choice of model on the basis of the observations. He will usually wish to make an estimate of the true response at each point of the factor space. Any estimate will be subject to variance of estimation due to error of observation, and to bias of estimation due to use of an improper model for fitting the observations. The aim of this thesis is to study bias and variance properties of

designs, to obtain, if possible, optimal designs with respect to certain criteria and the relative values of different designs.

We will consider first designs which minimize the maximal bias over the interval of interest when we fit a polynomial of degree $p$ and the true underlying model is a polynomial of degree $p+1$, and bias and variance properties of such designs will be studied in detail.

In many applications the experimenter prefers to use a low degree polynomial model even though a higher degree model seems to give a better fit. We will study the situation when this attitude is justified and determine designs which will give the experimenter an unbiased estimator of the average square deviation between the fitted polynomial and the assumed true underlying polynomial model of given degree. Box and Draper (1959) considered the study of the mean average square error, and obtained optimal designs with respect to the average square bias, and showed that for the fitting of polynomials of first degree those designs have good bias as well as variance properties. We shall consider the variance and bias properties of designs for the fitting of general one-dimensional polynomial of degree $p$. We shall determine optimal designs with respect to the average square bias weighted with a positive function and study the bias and variance properties of such designs. We shall compare the

efficiency of the optimal designs obtained and the one given in the literature by Guest (1958), Hoel (1958), and others, with respect to several measures related to bias and variance.

We shall then consider designs for estimating true yields on an n-dimensional lattice of points. Such a lattice of points is usually called a factorial system, and a design which consists of examining a subset of the points of the lattice is called fractional factorial design. We shall study two kinds of random fractional designs. Star designs will be defined and we will obtain the general form of the average mean square error for these designs in terms of bias and variance.

For the case of symmetrical factorial systems with the number of levels of each factor a prime or a power of a prime, a system for obtaining a subset of the totality of points is based on finite geometry. This is known as fractional replication and gives pleasing properties of balance. The bias and variance properties of such fractionation, in which a fraction is chosen at random, will be studied and the average mean square error over the totality of lattice points will be obtained. Finally the two systems for examining a lattice will be compared.

## B. Review of Literature

Since throughout this work we will refer extensively to orthogonal polynomials, for completeness we will briefly state some basic properties and definitions. These have been abstracted from Ghizzetti and Ossicini (1970), Sansone (1959), Isaacson and Keller (1966), and Todd (1963).

Definition. Let $w(x)$ be a non-negative function defined on the interval $[a,b]$ and such that all moments

$$\mu_n = \int_a^b x^n w(x)\,dx, \quad n = 0,1,2,\ldots$$

exist and are finite. A sequence of polynomials $p_n(x)$, of degree $n$, is said to be orthogonal with respect to the weight function $w(x)$ in $[a,b]$, if

$$\int_a^b p_m(x)p_n(x)w(x)\,d(x) = 0$$

for $m \neq n$.

Three basic properties are as follows:

1. We can always construct a sequence of polynomials $p_n(x)$ which are orthogonal with respect to $w(x)$.

2. All zeros of $p_n(x)$ are real, distinct, and lie in the interior of the interval $[a,b]$.

3. Denote by $\pi_n(x)$ the product $(x-x_1)\ldots(x-x_n)$.

The minimum of

$$\int_a^b \pi_n^2(x)w(x)\,dx$$

is achieved by letting $x_1,\ldots,x_n$ be the zeros of a polynomial orthogonal with respect to $w(x)$.

There is a very wide variety of polynomials determined by the interval $[a,b]$ and the weight function $w(x)$. We present a table with the classes of polynomials that arise in the present work.

We now review the design literature and we mention first the work concerned with the estimation aspect of polynomial regression.

G. Elfving (1952) considered the problem of allocation of observations to optimize the minimum variance unbiased estimator of a given linear function of the parameters. He planned to take $N$ observations at $R \leq N$ different points. He let $x_i$ be an allocation point and then consider the convex set determined by the vectors $x_i$, $i = 1,\ldots,R$, and its negatives.

It was then proved that the minimum variance unbiased estimator of a linear parametric function $\underline{\lambda}'\underline{\beta}$ is given by

$$\underline{\lambda}'\hat{\underline{\beta}} = \sum_{i=1}^{R} \ell_i \bar{y}_i$$

Table 1.1. Orthogonal polynomials

| Polynomial | Interval | Weight |
|---|---|---|
| Jacobi: $P_n^{(\alpha,\beta)}(x) = \frac{1}{2^n}\sum_{k=0}^{n}(-1)^{n-k}\binom{\alpha+n}{k}\binom{\beta+n}{k}(1-x)^{n-k}(1+x)^{k}$, $\alpha>0$, $\beta>0$ | $[-1,1]$ | $(1-x)^{\alpha-1}(1+x)^{\beta-1}$ |
| Chebyshev of 1st kind: $T_n(x) = \cos(n\cos^{-1}x)$ | $[-1,1]$ | $(1-x^2)^{-\frac{1}{2}}$ |
| Chebyshev of 2nd kind: $U_n(x) = \frac{1}{n+1}T_n'(x) = \frac{\sin\{(n+1)\cos^{-1}x\}}{\sin\{\cos^{-1}x\}}$ | $[-1,1]$ | $(1-x^2)^{\frac{1}{2}}$ |
| Legendre: $P_n(x) = \frac{(2n)!}{2^n(n!)^2}[x^n - \frac{n(n-1)}{2(2n-1)}x^{n-2} + \frac{n(n-1)(n-3)}{2x4(2n-1)(2n-3)}x^{n-4}$ ... + ...] | $[-1,1]$ | 1 |
| Laguerre: $L_n^{(\alpha)}(x) = \sum_{k=0}^{n}(-1)^k\binom{\alpha+n}{n-k}\frac{x^k}{k!}$, $\alpha>0$ | $[0,\infty)$ | $x^{\alpha-1}e^{-x}$ |
| Hermite: $H_n(x) = (2x)^n - \frac{n(n-1)}{1!} + \frac{n(n-1)(n-2)(n-3)}{2!}(2x)^{n-4} - \ldots$ | $(-\infty,\infty)$ | $e^{-x^2}$ |

with

$$\ell_i = \pm w_i|\lambda|/|v_\lambda|, \quad \sum_{i=1}^{R} w_i = 1$$

where $v_\lambda = \sum_{i=1}^{R} \pm w_i x_i$ is a vector on the convex set determined by the vectors corresponding to the different allocation points and in the same direction as $\underline{\lambda}$.

Commonly we will be interested not only in particular functions of the parameters but also in predicting responses at different points of the factor space. In any case, Elfving was one of the first to consider the problem of obtaining optimal designs in linear regression.

De La Garza (1954) was the first to work on the problem of relating allocation of observations to the generalized variance of estimates of parameters of a polynomial model. He assumed that the true underlying relationship was a polynomial of known degree $p$ and showed that the same information matrix, and hence the same generalized variance, can be obtained by replacing a given set of $N \geq p+1$, observations at the points $x_1, x_2, \ldots, x_N$ by $N$ observations made at only $p+1$ properly selected points in the interval from $x_1$ to $x_N$. De La Garza also did some work on a special case of the problem of minimization of the variance of the fitted polynomial inside the interval of experimentation.

Guest (1958), continuing De La Garza's work, considered

how to allocate the $N \geq p+1$ observations over the $[-1,+1]$ interval in order to minimize the maximal variance for prediction over the whole interval. He proved that if the degree of the true underlying polynomial model is $p$, then to minimize the maximal variance of the fitted curve over $[-1,+1]$ we should allocate an equal number of observations at the end points of the interval, as well as at the $p-1$ zeros of the derivative of the Legendre polynomial of degree $p$. The variance of a fitted point inside the interval when using such allocation was found to be

$$V[\hat{y}_p(x)] = \{1 + \frac{1-x^2}{p(p+1)} P'^2_p(x)\} \frac{(p+1)}{N} \sigma^2 \qquad (1.1)$$

where $P_p(x)$ is the Legendre polynomial of degree $p$. He considered polynomial fitting using equal spacing and obtained an approximate expression for the variance of a fitted polynomial in terms of Legendre polynomials given by

$$V[\hat{y}_p] \sim \{(p+1) P^2_p(x) - \frac{x^2-1}{(p+1)} P'^2_p(x)\}(p+1) \frac{\sigma^2}{N} \qquad (1.2)$$

Finally, he compared the behavior of the variance functions when fitting quadratic and cubic polynomials using the minimax variance designs and equal spacing of observations.

Hoel (1958) considered the generalized variance of the estimates as the criterion for optimization. He proved that when the degree of the true underlying polynomial is known, the allocation which is optimal using the generalized

variance criterion is the same one that Guest found optimal for the maximum of the variance criterion, i.e., equal number of observations at -1, +1 and at the zeros of the first derivative of a Legendre polynomial of the same degree as the true model. Using the same criterion, some results were obtained on the increase of efficiency arising from doubling the number of equally spaced observation points when the length of the interval is fixed and when it is doubled. Other measures of efficiency were also considered for polynomial models and for models based on a particular stationary process and a pure birth process.

Hoel and Levine (1964) considered allocation of observations to minimize the variance of a predicted value of the fitted polynomial at a specified point beyond the interval of experimentation. They also assumed that the degree of the true polynomial model was known before hand. Their main results are the following:

Lemma. If the $p_i$, $i = 0,1,\ldots,p$ are allowed to vary continuously in $(0,1)$ under the restriction $\sum_{i=0}^{p} p_i = 1$, then for a fixed $x \neq x_i$, $i = 0,1,\ldots,p$, the choice

$$p_i = \frac{|L_i(x)|}{\sum_{i=0}^{p} |L_i(x)|}, \quad i = 0,1,\ldots,p$$

where $L_i(x)$ is the Lagrange coefficient given by

$$L_i(x) = \frac{(x-x_0)(x-x_1)\ldots(x-x_{i-1})(x-t_{i+1})\ldots(x-x_p)}{(x_i-x_0)(x_i-x_1)\ldots(x_i-x_{i-1})(x_i-x_{i+1})\ldots(x_i-x_p)},$$

will minimize $V[\hat{y}(x)]$.

Theorem. If the minimizing p's given by the formula above are used, the p+1 observation points that will minimize $V[\hat{y}(x)]$ at $x>1$ are given by the Chebyshev points

$$x_i = -\cos\frac{i\pi}{p}, \quad i = 0,1,\ldots,p .$$

It was also prove that the preceding Chebyshev solution becomes a minimax variance solution for the interval $(-1,t)$, provided $t > t_1 > 1$ when $t_1$ is a value satisfying a certain equation. In later papers Hoel (1965a,b) extended the above work, determining minimum variance designs for two dimensional polynomial regression.

Kiefer and Wolfowitz (1960), and others unified and extended all the work on minimum variance and generalized variance mentioned above. They consider a polynomial regression model in which $E[\hat{y}(x)] = \sum_{i=0}^{p} \theta_i f_i(x)$ where $f_i(x)$, $i = 0,1,\ldots,p$ are real valued continuous functions, and $\theta = (\theta_1,\ldots,\theta_p)$ is an unknown parametric vector. An experimental design is regarded as specifying a probability measure concentrating mass $p_1,p_2,\ldots,p_r$ at the points $x_1,x_2,\ldots,x_r$ where the values $n_i = p_iN$, $i = 1,2,\ldots,r$ are integers. The associated experiment involves taking $n_i$ uncorrelated

observations on the random variable $y(x_i)$, $i = 1,2,\ldots,r$. Once a design is prescribed and the observations are made, the least squares method is used for estimating the parameters $\theta_0,\theta_1,\ldots,\theta_p$.

The problem then consists of deciding on an optimality criterion and finding a probability measure which is optimal for the criterion chosen.

Kiefer (1959) considered optimization of several different criteria:

1. M-optimality: Maximizing the infimum of power of test of a null hypothesis against a class of alternatives.
2. L-optimality: Maximizing the limiting power of test in the neighborhood of the null hypothesis.
3. E-optimality: Minimizing the maximum eigenvalue of the variance-covariance matrix of estimates, used by Wald (1943) and Ehrenfeld (1955).
4. A-optimality: Minimizing the trace of the variance-covariance matrix of estimates.
5. D-optimality: Minimizing the generalized variance of estimates of parameters.
6. G-optimality: Minimizing the maximum variance of prediction over the experimental region.

It should be noted that almost all the above criteria

are related only to the estimation aspect of the fitting problem and depend upon the assumption that the underlying model is known before experimentation.

In the following paragraphs we will continue presenting in detail only the results related to G-optimality and D-optimality.

As mentioned above, Guest (1958) and Hoel (1958) obtained the same optimal designs for the G-optimality and D-optimality criteria. Kiefer and Wolfowitz (1960) considered those problems under the probability measure approach and by using functional analysis proved that they were equivalent. Denoting by $\xi$ the probability measure which determines the design, $M(\xi)$ the information matrix, and $N^{-1}d(x,\xi)$ the variance function of the fitted curve, the main result of their paper is the following theorem:

The conditions

i. $\xi^*$ maximizes the determinant $M(\xi)$ ,

ii. $\xi^*$ minimizes $\max d(x,\xi)$,

iii. $\max d(x,\xi^*) = p+1$

are equivalent. The set B of all $\xi$ satisfying these conditions is convex and closed and $M(\xi)$ is the same for all $\xi \varepsilon B$.

Note here that this theorem is measure theoretic and has limited applicability because the design that achieves this optimality will consist of a probability distribution, say, $\xi(x_1), \xi(x_2), \ldots, \xi(x_n)$ which gives the proportion of observa-

vations at the points $x_1, x_2, \ldots, x_n$ respectively. If the problem is to allocate a pre-chosen definite number of observations, say $N$, the solution may not be achievable because $N\xi(x_i)$ will not in general be an integer.

Karlin and Studden (1966) considered experimental designs for the case where the functions $f_0, f_1, \ldots, f_p$ are of the form $\sqrt{w(x)}\, x^i$, $i=0,1,\ldots,p$ and $w(x) \geq 0$. For special choices of $w(x)$ the optimal designs (optimality is meant in the sense that $\xi$ maximizes $|M(\xi)|$) are obtained. When $w(x) = (1-x)^{\alpha+1}(1+x)^{\beta+1}$ $(\alpha > -1, \beta > -1)$ and $x$ is in the interval $[-1,+1]$, the optimal design $\xi$ consists of $p+1$ equal mass points located at the $p+1$ zeros of the Jacobi polynomial $P_{p+1}^{(\alpha,\beta)}(x)$.

Farrell, Kiefer and Walbran (1965) studied optimum multivariate designs under the same approach described above for the case where the experimental region $R$ is a compact subset of an Euclidean $k$-dimensional space. The main concern of their paper is to characterize, for several regions $R$ and functions $f = (f_0, \ldots, f_p)$ some $\xi$ which are D-optimum.

They treat in detail three cases:

i. $R = \{x;\ \sum_{i=1}^{k} x_i^2 \leq 1\}$

ii. $R = \{x;\ \max_{1 \leq i \leq k} |x_i| \leq 1\}$

iii. $R = \{x;\ x_i \geq 0,\ (i = 1,\ldots,k),\ \sum_{i=1}^{k} x_i = 1\}$.

All the work presented up to this point assumed that the model which explained the observations was known prior to experimentation. As a result, the several optimization criteria dealt only with estimation related aspects of polynomial regression. The model discrimination part of the problem, the one that refers to wrong model assumption and bias, received no attention.

Folks (1958) considered several criteria of optimality related to bias and variance in regression, like the minimization of the average mean square error, average mean square bias, maximal bias over a region, and so on. His results are limited to the fitting of polynomials of first degree.

Box and Draper (1959) considered the average mean square error over the region of interest R, as a basic criterion. This criterion involves both variance and bias.

They investigated the case where $\eta(\underline{x})$, a polynomial of degree t,

$$\eta_t(\underline{x}) = x_1'\underline{\beta}_1 + x_2'\underline{\beta}_2$$

is to be approximated by the fitted equation, a polynomial of degree s, with $s<t$, given by

$$\hat{y}_s(x) = \underline{x}_1'\underline{b}_1$$

where the vector $\underline{x}_1'$ is made up of the powers of the $x_i$'s in the polynomial of degree s, the vector $\underline{x}_2'$ is made up of the

additional higher order $x_i$'s terms in the polynomial of degree t, and $\underline{\beta}_1$ and $\underline{\beta}_2$ are the corresponding vectors of unknown coefficients. The estimator $\underline{b}_1$ is found by ordinary least squares to be

$$\underline{b}_1 = (X_1' X_1)^{-1} X_1' y$$

where $X_1$ is the matrix of values taken by the terms in $\underline{x}_1'$ at the locations of the independent variables and where $y$ is the column vector of the observations

$$y_i = \eta_t (\underline{x}_{1i}) + e_i, \quad i = 1,2,\ldots,N$$

where the $e_i$'s have zero mean and are uncorrelated with common variance $\sigma^2$.

The average mean square error over $R$ is

$$J = \Omega \int_R E[\hat{y}_s(x) - \eta_t(x)]^2 \, dx$$

where

$$\Omega^{-1} = \int_R dx.$$

$J$ can also be written as

$$J = B + V$$

where

$$B = \Omega \int_R E[\hat{y}_s(x) - \eta_t(x)]^2 \, dx$$

and

$$V = \Omega \int_R V[\hat{y}_s(x)]\, dx$$

Thus J is the sum of the average variance of the fitted polynomial, denoted by V, and the average square bias, denoted by B.

As expected, the minimization of J with respect to the choice of a design depends on the relative magnitudes of the B and V contributions. When they considered the special case of s=1 and t=2, the somewhat surprising conclusion of Box and Draper (1959) was that unless the V contribution was many times larger than the B contribution, optimum designs for the minimization of the bias component B alone were remarkably close to optimal designs for the case where the minimization was with respect to B+V. They showed that to minimize B alone, a design should satisfy the condition

$$(X_1' X_1)^{-1} X_1' X_2 = W_1^{-1} W_2$$

where $X_2$ is the matrix of values taken by the variables in $\underline{x}_2'$ (the omitted part of the model) and the matrices $W_1$ and $W_2$ are defined by

$$W_1 = \Omega \int_R \underline{x}_1 \underline{x}_1'\, dx ,$$

and

$$W_2 = \Omega \int_R \underline{x}_1 \underline{x}_2' \, dx.$$

These are the moment matrices of a uniform distribution over the region of interest.

We now consider the work on factorial experiments, i.e., the case in which the experimental variables are qualitative factors and the structure of the totality of possible information consists of the true yields and variability for each of the possible factor combinations. If one has factors say $F^{(1)}, F^{(2)}, \ldots, F^{(n)}$ with levels $x_1, x_2, \ldots, x_n$, the underlying formula for yield will be of the form

$$y(x_1, x_2, \ldots, x_n) = \eta(x_1, x_2, \ldots, x_n) + \text{error}$$

where the function is defined only for the factor levels $x_1, x_2, \ldots, x_n$. In other words, the model is classificatory. We will consider only the case where $\eta(x_1, x_2, \ldots, x_n)$ is a linear function of effects and interactions. There appears to be little general theory for non-linear classificatory models.

The most commonly used factorial plans involve factors which all occur with the same number of levels. Experiments with those plans are known as symmetrical factorial. The general theory of symmetrical factorial experiments has been considered in detail by Bose and Kishen (1940), Fisher (1943, 1945), and Bose (1947).

When a factorial experiment involves many factors, each of which is tested at several levels, it is well known that economy of space and material may be attained by observing only a fraction of all possible combination levels. Finney (1945) introduced the general idea of fractional replication. Kempthorne (1947) related ideas of fractional replication to confounding, and some basic aspects of general theory of fractional replication for prime power factorial systems are given by Kempthorne (1952). One of the earliest fractional replication plans to appear in the statistical literature was introduced by Tippett (1934). He utilized a 5 x 5 completely orthogonalized square to construct an orthogonal plan of six factors, each having five levels, with twenty-five treatment combinations.

Fisher (1945) using the properties of groups has given a system of confounding designs preserving main effects and all first order interactions.

Plans for two level factors, where all interactions are absent, were developed from considerations of weighting objects on a chemical balance scale. These problems were studied by Yates (1933) and Hotelling (1944); they obtained optimum weighing designs (in the sense of minimum variance) which yield orthogonal estimates of the weights.

Plackett and Burman (1946) obtained plans which permit estimation without correlation of all main effects of a

factorial arrangement, when the interactions are negligible. These plans are called orthogonal main-effect plans.

Rao (1946, 1947) generalized the ideas above introducing the concept of hypercubes of strength d. These plans lead to the orthogonal estimation of main effects and interactions up to the order k, when interactions of order greater than d($d > k$) are absent.

Bose and Bush (1952) presented upper bounds for the maximum possible number of factors that can be accommodated in orthogonal arrays of strength two and three. They also obtained methods for constructing these arrays. Addelman and Kempthorne (1961) have a general procedure which yields many orthogonal plans.

Methods for constructing fractional replicate plans in which all main effects are confounded with three-factors and higher order interactions are known as Resolution IV plans, and were developed by Box and Hunter (1961a). They also developed Resolution V plans (1961b), which allow orthogonal estimates of all two-factor interactions.

Frequently the Resolution V plans require more trials than one can afford to make. When this situation arises and some two-factor interactions are known to be important a compromise between main effect plans and Resolution V plans can be constructed. Three of such cases were discussed by Addelman (1962).

Chakravarti (1956) obtained fractional replicate plans for asymmetrical factorials arrangements. He constructed these plans by combining a main-effect plan for the $r^m$ arrangement with a main effect plan for the $s^m$ arrangement. These plans yielded orthogonal estimates of all main effects and all two-factor interactions which were comprised of one r-level factor and one s-level factor.

A catalogue of orthogonal main-effect plans with up to eighty-one treatment combinations is presented by Addelman and Kempthorne (1961).

It is not always possible to construct plans which permit orthogonal estimation of all important parameters with a small number of treatment combinations. Several procedures were developed for constructing non-orthogonal fractional replicate plans by Addelman (1961), and Whitwell and Morbey (1961).

Irregular fractions plans for $2^n$ arrangement were considered by Banerjee (1950) and Kempthorne (1952). Arrangements in which a subset of the factorial or a subset of the regular fractional factorial is duplicated were discussed by Daniel (1956) and further described by Dystra (1959). Addelman (1961) presented a general procedure for construction of $K/2^m$ replicate plans.

Satterthwaite (1959) was the first to consider the idea of examining factorial experiments using random sampling on

the choice of treatment combinations. Later, Dempster (1960, 1961) presented a theory of inference based on procedures similar to the ones given by Satterthwaite (1959).

Shah and Kempthorne (1962), have done initial work on the role of randomization in fractional factorials where the interactions assumed negligible are not so and the estimates of the parameters of interest are biased.

Ehrenfeld and Zacks (1967) established some procedures of testing hypothesis for random factorial experiments, which had been previously studied by them (Ehrenfeld and Zacks (1961, 1963)).

The common ANOVA F-tests for fractional replications of factorial experiments do not attain the assigned level of significance and power if the alias effects are not zero. Ehrenfeld and Zacks (1967) have shown that if the fractional replicates are properly randomized, then the usual F-like test statistics have asymptotically (for large number of nuisance parameters) the desired distribution.

Farrel (1968, 1969) studied the admissibility of randomized symmetrical designs. The problem is how to choose a design having a high power of the F-test, locally about the hypothesis. A partial theory of randomized designs is also developed from convexity considerations.

Lentner (1969) studied properties of generalized least squares estimators of a subgroup of pre-assigned parameters

of a $2^p$ factorial system when the design is a randomized fractional replicate of order $2^s (s<p)$ of the types suggested by Ehrenfeld and S. Zacks (1961). The class of generalized inverse solutions of the normal equations is characterized with respect to unbiasedness and optimality.

# II. MINIMAX BIAS DESIGNS

## A. Introduction

### 1. Aims

In this chapter we consider the problem of determining designs which will minimize the maximum of the $(\text{bias})^2$ function over the interval of interest when we fit a polynomial of degree $p$ and the true underlying functional relationship is a polynomial of degree $p + q$ with $q \geq 1$. Optimal designs will be obtained for the case where the fitted model has degree $p$ and the true model has degree $p + 1$. It is shown that for these designs the optimality remains valid if a model of degree $d < p$ is fitted and the true model has degree $d + 1$. The case of an underlying unknown nonlinear function is also considered and its relationship to the optimal designs mentioned above is established.

### 2. Illustration of the problem

Let us consider two special cases of the general problem. Since the linear transformation $z = [(b-a)x + (b+a)]/2$ establishes a one-to-one correspondence between any point $x$ in the $[-1, 1]$ interval and any $z$ in the general interval $[a,b]$, with $a$ and $b$ finite, without loss of generality, we let the region of interest be the interval $[-1, 1]$. We also present later a few results for the case of the extended intervals $[0, \infty)$ and $(-\infty, \infty)$.

First, suppose that the true relationship is a quadratic polynomial

$$\eta_Q(x) = \beta_0 + \beta_1 x + \beta_2 x^2 ,$$

while the model chosen for fitting is

$$y_{ij} = \beta_0 + \beta_1 x_i + e_{ij}$$

$$i = 1,2,\ldots,I, \quad j = 1,2,\ldots,n_i$$

where $n_i$ is the number of observations at the i-th location point, j is the j-th replication and the $e_{ij}$'s are uncorrelated random errors with expectation zero and common variance $\sigma^2$.

The properties of the least squares estimators are determined by the design matrix X in which rows relate the independent variable to the points of observation and the fitted model. It then follows that

$$X'X = N \begin{bmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{bmatrix} ,$$

where

$$\mu_k = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{n_i} n_i x_i^k , \quad k = 1,2,\ldots$$

is the general definition of design moments, and $N = \sum_{i=1}^{I} n_i$ is the total number of observations.

The least squares estimates of the parameters in the linear model are given by

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = \frac{1}{\mu_2 - \mu_1^2}\begin{bmatrix}\mu_2 & -\mu_1\\ -\mu_1 & 1\end{bmatrix}\begin{bmatrix}P_0\\ P_1\end{bmatrix}$$

where we set

$$P_k = \sum_{i=1}^{I}\sum_{j=1}^{n_i} n_i x_i^k y_{ij}, \quad k = 0,1,2,\ldots \ .$$

Under the assumption that the quadratic is the true model, these estimates are biased, since

$$E\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix} = \begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} + \frac{1}{\mu_2 - \mu_1^2}\begin{bmatrix}\mu_2^2 - \mu_1\mu_3\\ \mu_3 - \mu_1\mu_2\end{bmatrix}\beta_2 \ .$$

Hence the expected value of the fitted curve is

$$E\hat{y}_L(x) = (\beta_0 + \frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2}\beta_2) + (\beta_1 + \frac{\mu_3-\mu_1\mu_2}{\mu_2-\mu_1^2}\beta_2)\,x \ .$$

We define the $(\text{bias})^2$ function, resulting from fitting a polynomial of i-th degree when the true model has degree j as

$$B_{i,j}^2(x) = [E\hat{y}_i(x) - \eta_j(x)]^2 \ .$$

It follows from the above that

$$B^2_{1,2}(x) = \left[\frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2} + \frac{\mu_3-\mu_1\mu_2}{\mu_2-\mu_1^2}\, x - x^2\right]^2 \beta_2^2 \,. \qquad (2.1)$$

It seems intuitively reasonable that a good design will be symmetrical in the sense that all odd moments $\mu_1, \mu_3, \ldots$ will be zero. We consider this case so that the general expression given by 2.1 reduces to

$$B^2_{1,2}(x) = [\mu_2 - x^2]^2 \beta_2^2 \,. \qquad (2.2)$$

Let us find $\mu_2$ such that

$$\max_{-1\leq x\leq +1} (\mu_2 - x^2)^2$$

is a minimum.

If we put

$$f(x) = (a - x^2)^2$$

then

$$f'(x) = -4(ax - x^3), \quad f''(x) = -4(a - 3x^2),$$

so that

$$f'(x) = 0 \quad \text{for} \quad x = 0 \quad \text{or} \quad x^2 = a,$$

and

$$f''(0) = -4a, \quad f''(+\sqrt{a}) = +8a \,.$$

Since we cannot have a negative second moment, the case $a < 0$ will not be considered. For $a > 0$, $0$ is a point of relative maximum and the points $\pm \sqrt{a}$ are relative minima.

Note also that

$$f(0) = f(\pm \sqrt{2a}) = a^2 \quad \text{and} \quad f(\pm \sqrt{a}) = 0 .$$

For $x$ belonging to $[-\sqrt{2a}, +\sqrt{2a}]$ we will have $\max_x f(x) = a^2$ and for $x$ outside the above interval $\max_x f(x) > a^2$. Thus in order to minimize the maximum of $(a-x^2)^2$ in $[-1, 1]$ we should have $a = 1/2$.

Hence, if we restrict ourselves to symmetrical designs, in order to minimize the maximum of $B^2_{1,2}(x)$ over $[-1, 1]$, we should make $\mu_2 = 1/2$.

Since in the following we will make use of the generalized beta distribution, it seems appropriate to make some remarks about it at this point.

The beta distribution can be generalized to cover any finite interval. This leads to the following probability density function over the interval $[a,b]$, which is expressed in the notation of Hahn and Shapiro (1967)

$$f(x;\alpha,\beta;a,b) = \begin{cases} \frac{1}{a-b} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x-a}{a-b}\right)^{\alpha-1} \left(1 - \frac{x-b}{a-b}\right)^{\beta-1} \\ a \leq x \leq b, \quad \alpha > 0 \quad \text{and} \quad \beta > 0, \\ 0, \text{ elsewhere.} \end{cases}$$

In the following, the interval is usually $[-1, 1]$ and we

shall for brevity use the symbol $Be(\alpha,\beta)$ to refer to the distribution over this interval. We now consider two special cases of the distribution above to which we will refer quite often in this work.

When $\alpha = \beta = 1$ and $[a,b] = [-1, 1]$, the density function of this beta distribution reduces to

$$f(x;1,1;-1,1) = \begin{cases} 1/2, & -1 \leq x \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

This is the uniform or rectangular distribution over the interval $[-1, 1]$ and its moments are given by

$$\mu_{2n+1} = 0, \quad \mu_{2n} = \frac{1}{2n+1}, \quad n = 0,1,\ldots .$$

The other case is the one in which $\alpha = \beta = 1/2$ and $[a,b] = [-1, 1]$, the density function of this beta distribution is given by

$$f(x;1/2,1/2;-1,1) = \begin{cases} \frac{1}{\pi} \frac{1}{\sqrt{1-x^2}}, & -1 \leq x \leq 1 \\ 0, & \text{elsewhere} . \end{cases}$$

and its moments are

$$\mu_{2n+1} = 0, \quad \mu_{2n} = \frac{1}{2}\frac{3}{4}\cdots\frac{2n-1}{2n}, \quad n = 0,1,\ldots .$$

Note that the first two moments of the beta distribution $Be(1/2, 1/2)$ over $[-1, 1]$ are $\mu_1 = 0$ and $\mu_2 = 1/2$, and

that these are the design moments of the optimal design obtained above.

For the second special case, assume that the true underlying polynomial is a cubic

$$\eta_C(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

and the fitted model is a quadratic.

The X'X matrix for the fitted curve is

$$X'X = N \begin{bmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{bmatrix} .$$

It follows that the least squares estimates of the parameters are given by

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} \mu_2\mu_4-\mu_3^2 & \mu_2\mu_3-\mu_1\mu_4 & \mu_1\mu_3-\mu_2^2 \\ \mu_2\mu_3-\mu_1\mu_4 & \mu_4-\mu_2^2 & \mu_1\mu_2-\mu_3 \\ \mu_1\mu_3-\mu_2^2 & \mu_1\mu_2-\mu_3 & \mu_2-\mu_1^2 \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \\ P_1 \end{bmatrix} \qquad (2.3)$$

where

$$\Delta = \mu_2\mu_4 + 2\mu_1\mu_2\mu_3 - \mu_2^3 - \mu_3^2 - \mu_1^2\mu_4 .$$

Under the assumption that the cubic model is the true one, these estimates are biased, since

$$E\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix} + \frac{1}{\Delta}\begin{bmatrix}2\mu_2\mu_3\mu_4-\mu_3^3-\mu_1\mu_4^2-\mu_5\mu_2^2\\ \mu_2\mu_3^2-\mu_1\mu_3\mu_4+\mu_4^2-\mu_4\mu_2^2+\mu_1\mu_2\mu_5-\mu_3\mu_5\\ \mu_1\mu_3^2-\mu_3\mu_2^2+\mu_1\mu_2\mu_4-\mu_3\mu_4+\mu_2\mu_5-\mu_1^2\mu_5\end{bmatrix}\beta_3$$

where $\Delta$ is defined as above.

It follows that the general form of the $(\text{bias})^2$ function $B^2_{2,3}(x)$ is given by

$$B^2_{2,3}(x) = [\frac{1}{\Delta}(2\mu_2\mu_3\mu_4-\mu_3^3-\mu_1\mu_4^2-\mu_5\mu_2^2+\mu_1\mu_3\mu_5)$$

$$+ \frac{1}{\Delta}(\mu_2\mu_3^2-\mu_1\mu_3\mu_4+\mu_4^2-\mu_4\mu_2^2+\mu_1\mu_2\mu_5-\mu_3\mu_5)\ x$$

$$+ \frac{1}{\Delta}(\mu_1\mu_3^2-\mu_3\mu_2^2+\mu_1\mu_2\mu_4-\mu_3\mu_4+\mu_2\mu_5-\mu_1^2\mu_5)\ x^2-x^3]\ \beta_3^2$$

where

$$\Delta = \mu_4\mu_2+2\mu_1\mu_2\mu_3-\mu_2^3-\mu_3^2-\mu_1^2\mu_4 .$$

For symmetrical designs the above expression reduces to

$$B^2_{2,3}(x) = [(\mu_4/\mu_2)\ x-x^3]^2\ \beta_3^2 . \qquad (2.4)$$

Putting

$$f(x) = (ax - x^3)$$

we have

$$f'(x) = 0 \quad \text{for} \quad x = 0, \quad x = \pm\sqrt{a}, \quad x = \pm\sqrt{\frac{3a}{3}}$$

and

$$f''(0) = 2a, \quad f''(\pm\sqrt{a}) = \frac{7}{8}a^2, \quad f''(\pm\frac{\sqrt{3a}}{3}) = -8a^2 .$$

Also

$$f(0) = f(\pm\sqrt{a}) = 0$$

and

$$f(\pm\frac{\sqrt{3a}}{3}) = f(\pm\frac{2\sqrt{3a}}{3}) = \frac{4a^3}{27} .$$

It follows from the above that $x = \pm\frac{\sqrt{3a}}{3}$ are points of relative maxima and $x = 0$ and $x = \pm\sqrt{a}$ are points of relative minima.

Note that outside $[-\frac{2\sqrt{3a}}{3}, \frac{+2\sqrt{3a}}{3}]$ we will have $f(x) > \frac{4a^3}{27}$. Thus, the result follows making $\frac{2}{3}\sqrt{3a} = 1$ and $a = 3/4$.

Hence, if we restrict ourselves to symmetrical designs, to minimize the maximum of $B^2_{2,3}(x)$ over $[-1, 1]$ we should have

$$\mu_4/\mu_2 = 3/4.$$

In particular

$$\mu_2 = 1/2 \quad \text{and} \quad \mu_4 = 3/8$$

satisfy the above condition.

Note again that the moments of a beta distribution Be(1/2, 1/2) satisfy the particular conditions above.

## 3. Statement of the general problem

Suppose that the functional form that is chosen for fitting is a polynomial of degree p

$$\eta_p(x) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p$$

and the true underlying model is

$$\eta_{p+q}(x) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p + \ldots + \beta_{p+q} x^{p+q}$$

with $q \geq 1$.

The X'X matrix for the complete model can be partitioned as

$$X'X = N \left[\begin{array}{cccc:ccc} 1 & \mu_1 & \cdots & \mu_p & \mu_{p+1} & \cdots & \mu_{p+q} \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} & \mu_{p+2} & \cdots & \mu_{p+q+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \mu_p & \mu_{p+1} & & \mu_{2p} & \mu_{2p+1} & & \mu_{2p+q} \\ \hdashline \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \mu_{p+q} & \mu_{p+q+1} & \cdots & \mu_{2p+q} & \mu_{2p+q+1} & \cdots & \mu_{2(p+q)} \end{array}\right]$$

where the left upper submatrix is the X'X corresponding to the fitted p-th degree model.

The normal equations for the parameters in the fitted

model are given by

$$N \begin{bmatrix} 1 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \cdots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = \begin{bmatrix} P_0 \\ P_1 \\ \vdots \\ P_p \end{bmatrix} .$$

Under the assumption that the polynomial of degree $p+q$ is the true model, the estimates of the parameters in the fitted model are biased and their expected values are given by

$$E \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} 1 & \mu_1 & \cdots & \mu_p \\ \vdots & \vdots & \cdots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{bmatrix}^{-1} \begin{bmatrix} \mu_{p+1} & \cdots & \mu_{p+q} \\ \vdots & \cdots & \vdots \\ \mu_{2p+1} & \cdots & \mu_{2p+q} \end{bmatrix} \begin{bmatrix} \beta_{p+1} \\ \vdots \\ \beta_{p+q} \end{bmatrix}$$

or

$$E \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \hat{\beta}_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \Delta_{00} & \Delta_{01} & \cdots & \Delta_{0q} \\ \Delta_{10} & \Delta_{11} & \cdots & \Delta_{1q} \\ \vdots & \vdots & & \vdots \\ \Delta_{p0} & \Delta_{p1} & \cdots & \Delta_{pq} \end{bmatrix} \begin{bmatrix} \beta_{p+1} \\ \cdot \\ \cdot \\ \beta_{p+q} \end{bmatrix}$$

where $\Delta_{ij} = \Delta_{ij}(\mu_1, \ldots, \mu_{2p+q})$ for $i = 0,1,\ldots, p$ and $j = 0,1,\ldots,q$.

Hence the $(\text{bias})^2$ due to fitting a polynomial of

degree $p$ when the true model has degree $p + q$, $q \geq 1$, can be expressed as

$$B^2_{p,p+q}(x) = [(\Delta_{00}+\Delta_{10}x +\ldots+ \Delta_{p0}x^p - x^{p+1})\beta_{p+1}(\Delta_{01} + \Delta_{11}x +\ldots+ \Delta_{p1}x^p - x^{p+2})\beta_{p+2} +\ldots + (\Delta_{0q}+\Delta_{1q}x +\ldots+ \Delta_{pq}x^p - x^{p+q})\beta_{p+q}]^2 . \qquad (2.5)$$

It is the aim of this chapter to study the conditions under which one can obtain designs which will minimize the maximum of $B^2_{p,p+q}(x)$ over the interval of interest.

## B. Characterization of the Minimax Bias Designs

### 1. The Chebyshev allocation

We will study first in detail the case in which the fitted model is a polynomial of degree $p$ and the true underlying functional relationship is a polynomial of degree $p+1$.

The statistical expression of the observations in terms of the fitted model is given by

$$y_{ij} = \beta_0 + \beta_1 x_i +\ldots+ \beta_p x_i^p + e_{ij},$$

$i = 1,2,\ldots,p+1$, and $j = 1,2,\ldots,n_i$, where the $e_{ij}$'s are uncorrelated random errors each with mean zero and variance $\sigma^2$.

If we let $\bar{y}_i$ be the mean of the observations at $x_i$, then the least squares fit is determined completely by $\bar{y}_1,\ldots,\bar{y}_{p+1}$ and is the unique polynomial of degree $p$ which passes through these points. This curve can be expressed in terms of a Lagrange interpolation polynomial as

$$\hat{y}(x) = \sum_{i=1}^{p+1} L_i(x)\bar{y}_i ,$$

where

$$L_i(x) = \frac{(x-x_1)(x-x_2)\ldots(x-x_{i-1})(x-x_{i+1})\ldots(x-x_{p+1})}{(x_i-x_1)(x_i-x_2)\ldots(x_i-x_{i-1})(x_i-x_{i-1})\ldots(x_i-x_{p+1})} .$$

It follows that the expected value of the fitted curve is the interpolation polynomial through the points with coordinates $(x_i, \eta_{p+1}(x_i))$, for $i = 1,2,\ldots,p+1$; hence from the expression of the error in polynomial interpolation (Isaacson and Keller, 1966), the $(\text{bias})^2$ function is

$$B^2_{p,p+1}(x) = \pi^2_{p+1}(x)\beta^2_{p+1} ,$$

where

$$\pi_p(x) = (x-x_1)(x-x_2)\ldots(x-x_{p+1}).$$

For the case under consideration from 2.5 we can also write the $(\text{bias})^2$ function as

$$B^2_{p,p+1}(x) = [\Delta_0 + \Delta_1 x +\ldots+ \Delta_p x^p - x^{p+1}]^2\beta^2_{p+1}$$

where we put $\Delta_i \equiv \Delta_{i0}$, $i = 0,1,\ldots,p$ .

Since of all polynomials of degree $p+1$ having a leading coefficient equal to 1 the polynomial $2^{-p} T_{p+1}(x)$, has the smallest least-upper bound for its absolute value in the interval $[-1,1]$, it follows that the maximum of $B^2_{p,p+1}(x)$ will be minimum if and only if $B_{p,p+1}(x) = 2^{-p} T_{p+1}(x)$. Note that from the expression for $B^2_{p,p+1}(x)$ when we fit a $p$-th degree polynomial using only $p+1$ allocation points it follows that the condition above will be satisfied if we take the allocation points $x_1, x_2, \ldots, x_{p+1}$ equal to the zeros of $T_{p+1}(x)$. Thus we are led to the following result.

Theorem 2.1. Suppose a polynomial of degree $p$ is fitted and let the true underlying model be a polynomial of degree $p+1$. The necessary and sufficient condition for a design to minimize the maximum of the $(\text{bias})^2$ function over the interval $[-1,+1]$ is to have $B_{p,p+1}(x)$ equal to $2^{-p} T_{p+1}(x)$, where $T_{p+1}(x)$ is the Chebyshev polynomial of degree $p+1$. The design which places the observations at the zeros of $T_{p+1}(x)$ satisfy the above condition.

The allocation with an unequal number of observations at the Chebyshev zeros is optimal only for the case considered above. It will not be optimal if we fit a polynomial of degree $d < p$ and the true model has degree $d + 1$. As we will see in the following the optimal property of the Chebyshev allocation will hold through sequential fitting only when we have have an equal number of observations at each of the Chebyshev zeros.

## 2. The beta distribution Be(1/2, 1/2) moment conditions

The general expression of the Chebyshev quadrature formula (Ghizetti and Ossicini, 1970) is

$$\int_{-1}^{+1} \frac{f(x)}{\sqrt{1-x^2}}\,dx = \frac{\pi}{p+1}\sum_{i=1}^{p+1} f(x_i) + \frac{2\pi}{(p+2)2^{p+1}} f(\xi)^{(2p+2)} \qquad (2.6)$$

where

$$|\xi| < 1$$

and

$$x_i = \cos\left(\frac{2i-1}{p+1}\,\frac{\pi}{2}\right), \quad i = 1,2,\ldots,p+1$$

are the zeros of the Chebyshev polynomial of degree $p+1$ over $[-1, 1]$.

Putting $f(x) = x^k$, in 2.6, for $k = 1,2,\ldots,2p+1$ we will have

$$\frac{1}{\pi}\int_{-1}^{+1} \frac{x^k}{\sqrt{1-x^2}}\,dx = \frac{1}{p+1}\sum_{i=1}^{p+1} x_i^k \,. \qquad (2.7)$$

The left-hand side of the last equality is the $k$-th moment of the beta distribution Be(1/2, 1/2) over $[-1, 1]$. It follows that a design which consists of an equal number of observations at each of the zeros of the Chebyshev polynomial of degree $p+1$ has its design moments up to the order $2p+1$ equal to the moments of the beta distribution under consideration.

According to Theorem 2.1, the allocation above minimizes the maximum of $B^2_{p,p+1}(x)$ over $[-1, 1]$. Also as noted above for our case 2.5 reduces to

$$B^2_{p,p+1}(x) = [\Delta_0+\Delta_1 x +\ldots+ \Delta_p x^p - x^{p+1}]^2 \beta^2_{p+1} \qquad (2.8)$$

where the bias terms $\Delta_i \equiv \Delta_{i0}$, $i = 0,\ldots,p$, now depend on the design moments up to order $2p + 1$, or, $\Delta_i = \Delta_i(\mu_1,\mu_2,\ldots,\mu_{2p+1})$. Hence it follows that the design moments, which minimize the maximum of $B^2_{p,p+1}(x)$, over $[-1, 1]$ are the first $2p + 1$ moments of the beta distribution $Be(1/2, 1/2)$. Note that using the appropriate denominator in the right-hand side 2.7 holds for the zeros of any Chebyshev polynomial of degree $d \geq p+1$. We can now state the following:

Theorem 2.2. Suppose a polynomial of degree $p$ is fitted and the true underlying model has degree $p + 1$. A sufficient condition to minimize the maximum of the $(\text{bias})^2$ function over the interval of interest is to have the design moments up to the order $2p + 1$ equal to the corresponding moments of the beta distribution $Be(1/2, 1/2)$ defined in the same interval. In particular, the above condition is satisfied by allocating an equal number of observations at the zeros of a Chebyshev polynomial of degree $n$, with $n \geq p + 1$.

Note that an attractive feature of these designs is that the optimality remains valid if a model of degree $d < p$ is fitted and the true model is of degree $d + 1$.

The designs which are optimal with respect to the minimax

bias criterion will be called "minimax bias" designs throughout the remainder of this thesis.

## 3. Necessary and sufficient conditions for minimax bias designs for fitting linear and quadratic polynomial models

According to 2.8 we can write

$$B^2_{p,p+1}(x) = [\Delta_0 + \Delta_1 x + \ldots + \Delta_p x^p - x^{p+1}]^2 \beta^2_{p+1}$$

where

$$\Delta_i = \Delta_i(\mu_1, \ldots, \mu_{2p+1}) \quad \text{for} \quad i = 0,1,\ldots,p.$$

It follows from Theorem 2.2 that if we make the $\mu_j$'s in the expression for the $\Delta_i$'s equal to the corresponding moments of the beta distribution $Be(1/2, 1/2)$ we will obtain the $\tilde{\Delta}_i$'s values which will minimize the maximum of $B^2_{p,p+1}(x)$.

Thus, if we solve the system of equations in the $\mu_i$'s

$$\Delta_i(\mu_1, \ldots, \mu_{2p+1}) = \tilde{\Delta}_i \ , \quad 0 = 1,2,\ldots,p$$

we will obtain the necessary and sufficient design moments conditions for the problem under consideration.

When we fit a linear polynomial and the true model is a quadratic polynomial, according to 2.1, the $(\text{bias})^2$ function is of the form

$$B^2_{1,2}(x) = [\Delta_0 + \Delta_1 x - x^2]^2 \ \beta^2_2$$

where

$$\Delta_0 = \frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2} \quad \text{and} \quad \Delta_1 = \frac{\mu_3 - \mu_1\mu_2}{\mu_2 - \mu_1^2} .$$

If we substitute the moments of the beta distribution Be(1/2, 1/2) in the above expression, we will have

$$\Delta_0 = \frac{1}{2} \quad \text{and} \quad \Delta_1 = 0.$$

Hence we have the nonlinear system

$$\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2} = \frac{1}{2} \quad \text{and} \quad \frac{\mu_3 - \mu_1\mu_2}{\mu_2 - \mu_1^2} = 0 ,$$

which yields

$$\mu_2 = \frac{1}{2} \quad \text{and} \quad \mu_3 = \frac{\mu_1}{2} . \tag{2.9}$$

We summarize the above as follows:

Theorem 2.3. The necessary and sufficient design moment conditions for the maximum of the $(\text{bias})^2$ function to be minimum over the [-1, 1] interval when a first degree polynomial is fitted and the true polynomial model is a quadratic polynomial are given by

$$\mu_2 = \frac{1}{2} \quad \text{and} \quad \mu_3 = \mu_1/2 ,$$

with $\mu_2 \neq \mu_1^2$. Note that we will obtain the moments of a beta Be(1/2, 1/2) distribution over [-1, 1] if we let

$\mu_1 = 0$.

Let us now consider the case where the model fitted is a quadratic and the true model is a cubic.

$$B^2_{2,3}(x) = [\Delta_0 + \Delta_1 x + \Delta_2 x^2 - x^3]\ \beta^2_3$$

where

$$\Delta_0 = [2\mu_2\mu_3\mu_4 - \mu_3^3 - \mu_1\mu_4^2 - \mu_5\mu_2^2 + \mu_1\mu_3\mu_5)/\Delta$$

$$\Delta_1 = (\mu_2\mu_3^2 - \mu_1\mu_3\mu_4 + \mu_4^2 - \mu_4\mu_2^2 + \mu_1\mu_2\mu_5 - \mu_3\mu_5)/\Delta$$

$$\Delta_2 = (\mu_1\mu_3^2 - \mu_3\mu_2^2 + \mu_1\mu_2\mu_4 - \mu_3\mu_4 + \mu_2\mu_5 - \mu_1^2\mu_5)/\Delta$$

and

$$\Delta = \mu_4\mu_2 + 2\mu_1\mu_2\mu_3 - \mu_2^3 - \mu_3^2 - \mu_1^2\mu_4 .$$

Substituting the moments of a beta distribution Be(1/2, 1/2) in the expressions above we will have

$$\Delta_0(\mu_1,\ldots,\mu_5) = 0, \quad \Delta_1(\mu_1,\ldots,\mu_5) = \frac{3}{4} ,$$

and

$$\Delta_2(\mu_1,\ldots,\mu_5) = 0 .$$

For purposes of sequential fitting of polynomials of increasing degree we are interested in minimax bias designs that will also be optimal if we fit a linear polynomial and the true polynomial is a quadratic, so let us put $\mu_2 = 1/2$ and $\mu_3 = \mu_1/2$ in the system above, and solve for

the remaining parameters. We will then have the following solution

$$\mu_4 = \frac{3}{8} \quad \text{and} \quad \mu_5 = 0 \ .$$

We state the above as follows:

Theorem 2.4. The necessary and sufficient design moment conditions for the maximum of the $(\text{bias})^2$ function to be minimum over the $[-1, 1]$ interval when a linear polynomial model is fitted and the true model is a quadratic polynomial, or when the fitted model is a quadratic polynomial and the true model is a cubic polynomial is given by

$$\mu_1 = 0, \quad \mu_2 = 1/2, \quad \mu_3 = 0, \quad \mu_4 = 3/8, \quad \text{and} \quad \mu_5 = 0 \ . \tag{2.10}$$

Note that the design moment values above are the ones of a beta $\text{Be}(1/2, 1/2)$ distribution over the $[-1, 1]$ interval.

## 4. Properties of the minimax bias design

Because of the design moment properties of the Chebyshev allocation given in Theorem 2.2, any minimax bias design is such that

$$B^2_{p,p+1}(x) = 4^{-p} T^2_{p+1}(x) \beta^2_{p+1} \ . \tag{2.11}$$

Another property of the Chebyshev polynomial (Todd, 1963), is that

$$\int_{-1}^{+1} T_j^2(x)\,dx = 1 - \frac{1}{4j^2-1}\,, \quad j = 0,1,\ldots, \tag{2.12}$$

so substituting in 2.11 we have

$$1/2 \int_{-1}^{+1} B_{p,p+1}^2(x)\,dx = 4^{-p}\,\frac{2p^2+4p+1}{4p^2+8p+3}\,\beta_{p+1}^2\,.$$

Also note that the minimum of the modulus of $T_{p+1}(x)$ is attained at its zeros and that its maximum value 1 is attained at the points

$$x_i = \cos\left(\frac{i\pi}{p+1}\right)\,, \quad i = 0,1,\ldots,p+1, \tag{2.13}$$

(Isaacson and Keller, 1966). Hence we can state the following theorem:

Theorem 2.5. For any minimax bias design, the $(\text{bias})^2$ function and the mean average square bias, respectively, are given by

$$B_{p,p+1}^2(x) = 4^{-p}T_{p+1}^2(x)\beta_{p+1}^2$$

and

$$B_{p,p+1} = 4^{-p}\,\frac{2p^2+4p+1}{4p^2+8p+3}\,\beta_{p+1}^2\,.$$

The points of zero bias are the zeros of $T_{p+1}(x)$ and the maximum $(\text{bias})^2$ $4^{-p}\beta_{p+1}^2$ is attained at the points $x_i = \cos\left(\frac{i\pi}{p+1}\right)$, $i = 0,1,\ldots,p+1$.

From Table 1.1 the Chebyshev polynomial of degree $p + 1$, $T_{p+1}(x)$ is defined as

$$T_{p+1}(x) = \cos[(p+1)\cos^{-1}x], \quad p = 0,1,\ldots$$

and as noted before, its zeros are given by

$$x_i = \cos\theta_i, \quad \text{where } \theta_i = \frac{2i-1}{p+1}\frac{\pi}{2}, \quad i = 1,2,\ldots,p+1 .$$

We also have (Hildebrand, 1956),

$$\sum_{i=1}^{p+1} T_j(x_i)\, T_k(x_i) = \begin{cases} 0 & (j \neq k) \\ \frac{p+1}{2} & (j = k \neq 0) \\ p+1 & (j = k = 0) \end{cases}, \qquad (2.14)$$

where $j$ and $k$ are nonnegative integers not exceeding $p + 1$, so the Chebyshev polynomials $T_0(x), T_1(x),\ldots,T_p(x)$ are orthogonal under summation over the zeros of $T_r(x)$, with $r \geq p + 1$.

The above facts lead to the well-known fact that the Chebyshev zeros provide orthogonal least squares when we fit a p-th degree polynomial

$$\eta_p(x) = \sum_{j=0}^{p} c_j T_j(x),$$

taking an equal number of observations at the zeros of $T_p(x)$

with $r \geq p + 1$. The property just mentioned allows us to obtain a general form for the maximum of the variance function of a p-th degree polynomial, fitted using a minimax bias design.

The variance function of a fitted polynomial depends on the design only through its corresponding design moments. Hence, without loss of generality, the variance function of a p-th degree fit using a minimax bias design is the same as the variance function of the orthogonal fit mentioned above. Thus taking n observations at each zero of $T_{p+1}(x)$ we can write

$$V[\hat{y}_p(x)] = \sum_{j=0}^{p} \left\{ \frac{T_j^2(x)}{\sum_{i=1}^{p+1} T_j^2(x_i)} \right\} \frac{\sigma^2}{n} , \tag{2.15}$$

and since $N = n(p+1)$ it follows that

$$V[\hat{y}_p(x)] = \left\{ 1 + (p+1) \sum_{j=1}^{p} \frac{T_j^2(x)}{\sum_{i=1}^{p+1} T_j^2(x_i)} \right\} \frac{\sigma^2}{N} .$$

From the orthogonality property given in 2.14, we can write the above as

$$V[\hat{y}_p(x)] = \left\{ 1 + 2 \sum_{j=1}^{p} T_j^2(x) \right\} \frac{\sigma^2}{N} . \tag{2.16}$$

Since $|T_j(x)| \leq 1$, $j = 0,1,2,\ldots,$ and, as given in 2.13,

$|T_j(x)| = 1$ at the points $\xi_i = \cos(\frac{i\pi}{j+1})$, $i = 0,1,\ldots,j+1$, it follows that $-1$ and $1$ are the only common points where $T_1(x)$, $T_2(x)$,..., assume maximum modulo. It is therefore easy to see from 2.16 that the following result holds.

Theorem 2.6. If a p-th degree polynomial is fitted using a minimax bias design the maximum of its variance function is equal to $(2p+1)\sigma^2/N$ and that value is attained at the points $+1$ and $-1$.

According with 2.12 we have

$$\int_{-1}^{+1} T_j^2(x)dx = 1 - \frac{1}{4j^2-1}, \quad j = 0,1,\ldots, \tag{2.17}$$

and it can be proved by induction that

$$\sum_{j=1}^{p} \frac{1}{4j^2-1} = \frac{p}{2p+1}. \tag{2.18}$$

Since from 2.15

$$V_p = 1/2 \int_{-1}^{1} V[\hat{y}_p(x)]dx$$

$$= 1 + \sum_{j=1}^{p} \int_{-1}^{+1} T_j^2(x)dx,$$

it follows from 2.17 and 2.18 that

$$V_p = \frac{2p^2+2p+1}{2p+1} \frac{\sigma^2}{N}.$$

We can summarize the above result as follows.

Theorem 2.7. If a p-th degree polynomial is fitted using a minimax bias design the mean average variance is given by

$$V_p = [(2p^2+2p-1)/(2p+1)]\sigma^2/N \ .$$

## C. Considerations of the General Problem

### 1. The existence of an optimal design when the difference between the degree of the fitted and true polynomial is greater than 1

Let the fitted linear polynomial be

$$\hat{y}_L(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

and assume that the true underlying model is a cubic polynomial given by

$$\eta_C(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \ .$$

Following the general procedure presented in Section A.3 of this chapter

$$E\hat{y}_L(x) - \eta_C(x) = \left[\frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2} + \frac{\mu_3-\mu_1\mu_2}{\mu_2-\mu_1^2}\, x-x^2\right]\beta_2 + \left[\frac{\mu_2\mu_3-\mu_1\mu_4}{\mu_2-\mu_1^2} + \frac{\mu_4-\mu_1\mu_3}{\mu_2-\mu_1^2}\, x-x^3\right]\beta_3 \ , \qquad (2.19)$$

and for any minimax bias design 2.19 reduces to:

$$E\hat{y}_L(x) - \eta_C(x) = [\,1/2-x^2]\beta_2 + [(3/4)x-x^3]\beta_3 \,. \qquad (2.20)$$

It follows from 2.20 that for $\beta_2 \neq 0$ and $\beta_3 = 0$ we have

$$E\hat{y}_L(x) - \eta_C(x) = [1/2-x^2]\beta_2 \,, \qquad (2.21)$$

and -1,0 and +1, are the only three points at which the difference 2.21 assumes its maximum absolute value with alternating signs.

For $\beta_2 = 0$ and $\beta_3 \neq 0$ the difference 2.20 reduces to

$$E\hat{y}_L(x) - \eta_C(x) = [(3/4)x-x^3]\beta_3 \,, \qquad (2.22)$$

and -1, -1/2, +1/2, and +1 are the only four points at which the difference 2.22 assumes its maximum absolute value with alternating signs.

It follows from the above and Chebyshev's Theorem (Isaacson and Keller, 1966) that the minimax approximation for $\eta_C(x)$ depends on the values of its parameters and so there is no design which minimizes $[E\hat{y}_L(x) - \eta_C(x)]^2$ independently of $\beta_2$ and $\beta_3$ over the [-1, 1] interval.

The relationship between the result above and the general problem mentioned in Section A.3 of this chapter is immediate, and can be stated as follows. Let the fitted model be a polynomial of degree p and the true underlying polynomial model have degree p+q with q>1. There exists no design which minimizes the $(\text{bias})^2$ function, $B^2_{p,p+q}(x)$, q>1, over the interval of interest independently of the values of the

omitted parameters.

In summary, there exists no optimal design for the maximum bias criterion if the difference between the degree of the fitted and true polynomial models is greater than 1. In the next section we consider approximate solutions for this more general case.

## 2. Optimal allocation for non-linear models

We now consider the general case where the true underlying model is an unknown function, $f(x)$, and then apply our results to the case where the fitted polynomial has degree $p$ and the true model has degree $p + q$ with $q > 1$.

Under the assumption above, let the observational model be

$$y_i = f(x_i) + e_i$$

where the errors of observation $e_i$'s have expectation zero, common variance $\sigma^2$, and are uncorrelated.

Suppose that a polynomial of degree $p$ is fitted taking an arbitrary number of observations at $p + 1$ different values of $x$: $x_1, x_2, \ldots, x_{p+1}$. As in the derivation of Theorem 1.1, the expected fitted polynomial is the interpolation polynomial of degree $p$ passing through the points with coordinates $(x_i, f(x_i))$ for $1 = 1,2,\ldots,p+1$. Hence (Isaacson and Keller, 1966),

$$B^2_{p,f}(x) = \pi^2_{p+1}(x) \left[\frac{f(\xi)^{(p+1)}}{(n+1)!}\right]^2$$

where

$\min(x,x_1,\ldots,x_{p+1}) < \xi < \max(x,x_1,\ldots,x_{p+1})$, $f(x)$ is $p+1$ times differentiable, and $B^2_{p,f}(x)$ stands for the $(\text{bias})^2$ function resulting from fitting a polynomial of degree $p$ where the true model is given by $f(x)$.

Since $\xi = \xi(x,x_1,\ldots,x_{p+1})$, we cannot be sure that the Chebyshev zeros will minimize the maximum of $B^2_{p,f}(x)$. On the other hand, if $M_{p+1}$ is the maximum of

$$\left[f(x)^{(p+1)}\right]$$

we can minimize the quantity

$$\frac{\max B^2_{p,f}(x)}{M_{p+1}} = \frac{\max \pi^2_{p+1}(x)}{(p+1)!}$$

by choosing $x_1,\ldots,x_{p+1}$ to be zeros of $T_{p+1}(x)$.

We have then insured that

$$B^2_{p,f}(x) \leq \frac{M_{p+1}}{2^p(p+1)!} .$$

Moreover, if $M_{p+1}$ is finite for all $p$ the $(\text{bias})^2$ function $B^2_{p,f}(x) \to 0$ as $p \to \infty$.

It follows that for analytic functions $f(x)$ we can approximate as closely as desired, at all points in $[-1, 1]$, by choosing polynomials of successfully higher degrees whose

expected values agree with $f(x)$ at the zeros of the relevant Chebyshev polynomials.

In fact, the following theorem (Isaacson and Keller, 1966), shows the appropriateness of the Chebyshev allocation for least squares fit relaxing the analyticity condition mentioned above.

Theorem. Let $f(x)$ have continuous second derivative on $[-1, 1]$. If $P_{p,k}(x)$ is the discrete least squares approximation of degree at most $p$ to $f(x)$, with $\sqrt{p} \leq k \leq p$, based on the $p + 1$ points

$$x_i = \cos\left(\frac{2i-1}{p+1}\,\frac{\pi}{2}\right), \quad i = 1,2,\ldots,p+1$$

then $P_{p,k}(x)$ converges to $f(x)$ on $[-1, 1]$ as $p \to \infty$.

The above results are particularly interesting since for other selections of the $p+1$ points of observation, the bias may not tend to zero as $p$ increases. For equal spacing for example, and the function $(1+25x^2)^{-1}$ the bias increases without bound, at all except the points of observation, in the approximate range $0.7 \leq |x| \leq 1.0$, (Fox and Parker, 1968).

It is important to notice that the comments above hold for the case where $f(x)$ is a polynomial of degree $p + q$ with $q \geq 1$. Note that since any equal allocation of observations at the zeros of $T_{p+1}(x)$ has its first $2p+1$ moments

equal to the corresponding moments of the beta distribution Be(1/2, 1/2), it follows that minimax bias designs have the properties described above.

# III. MIN-AVERAGE BIAS DESIGNS

## A. Introduction

### 1. Aims

There are many situations in which we have enough data to fit a polynomial of high degree but the fit of a low degree model is preferred. In most applications, a lower degree model is easier to interpret. If the lower degree model predicts reasonably well it may be considered more appropriate than a more sophisticated model. However, even though the fitting of a low degree model is preferred, an essential condition is that on the average the model fitted should deviate little from the true model and an estimate of such deviation be available.

Designs which will minimize the average mean square bias and allow its estimation are useful to the experimenter if he needs to fit a low degree model when the true model has a higher degree. The determination of designs which satisfy both of the above conditions is one of the aims of the work in this chapter.

We will generalize the criterion of optimalization considered by Box and Draper (1959), considering optimal designs with respect to the average square bias weighted with a general class of functions. The particular cases where the optimal allocations are at the zeros of classical orthogonal polynomials are studied. Special attention is given to the

case of the Chebyshev weight function, and it is shown that the corresponding class of optimal designs and the minimax bias designs are the same.

## 2. Statement of the problem

The previous chapter was concerned with one-dimensional designs. In the following section both one-dimensional and multi-dimensional problems are considered. We therefore give the necessary multi-dimensional formulation, which was originally presented by Box and Draper (1959). We then state the specific problem to be considered in this section.

Let us assume that the true response surface is an unknown polynomial whose degree and form we want to ascertain from the observations $y_1,\ldots,y_N$ taken at properly chosen allocation vectors $\underline{x}_1,\ldots,\underline{x}_N$, and suppose that the true model is a polynomial of degree $s$ given by

$$\eta_s(\underline{x}) = \underline{x}_1' \underline{\beta}_1 \tag{3.1}$$

where

$$\underline{x}_1' = (1, x_1,\ldots,x_k;\ x_1^2,\ldots,x_k^2;\ x_1x_2,\ldots)$$

$$\underline{\beta}_1' = (\beta_0,\beta_1,\ldots,\beta_k,\ \beta_{11},\ldots,\beta_{kk},\ \beta_{12},\ldots),$$

contain terms up to the order $s$.

With observations at $N$ points we can define the design matrix $X_1$, and denoting by $y$ the observed values at those points we will have

$$y = X_1\underline{\beta}_1 + \underline{e}$$

$$X_1 = \begin{bmatrix} \underline{x}'_{11} \\ \vdots \\ \underline{x}'_{1u} \\ \vdots \\ \underline{x}'_{1N} \end{bmatrix}$$

where the column vectors of $X_1$ are assumed to be linearly independent and the column vector of the errors has expectation zero and variance - covariance $I_N\sigma^2$.

If the assumed polynomial model given by 3.1 is the true one, the best linear unbiased estimator $\underline{b}_1$ of $\underline{\beta}_1$ is given by the least squares estimate

$$\underline{b}_1 = (X_1'X_1)^{-1} X_1'y \tag{3.2}$$

and its variance-covariance matrix is given by

$$E(\underline{b}_1-\underline{\beta}_1)\ (\underline{b}_1-\underline{\beta}_1)' = (X_1'X_1)^{-1}\sigma^2 .$$

The fitted polynomial is thus

$$\hat{y} = X_1\underline{b}_1 .$$

Suppose that the model assumed in 3.1 is not correct, and that additional terms are needed, so the true underlying polynomial model has degree $t > s$, or

$$\eta_t\ (x) = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2 . \tag{3.3}$$

The corresponding X matrix is now given by

$$X = (X_1, X_2)$$

with $X_2$ and $\underline{\beta}_2$ similar to $X_1$ and $\underline{\beta}_1$ defined above, but with terms from order $s+1$ up to $t$.

Now the least squares estimate given by 3.2 is biased for

$$E(\underline{b}_1) = \underline{\beta}_1 + A\underline{\beta}_2$$

where $A = (X_1'X_1)^{-1} X_1'X_2$ is the so-called "alias" matrix. The residual sum of squares will now also be a biased estimator of $(N-s)\sigma^2$, and we will have

$$E(y'y-\underline{b}_1'(X_1'X_1)^{-1}\underline{b}_1) = (N-s)\sigma^2+N\underline{\beta}_2'(X_2'X_2-X_2'X_1(X_1'X_1)^{-1}X_1'X_2)\underline{\beta}_2 .$$

We will consider as the overall measure of the goodness of the design the well known average mean square error

$$J = [\int_R [E\hat{y}_s-\eta_t(x)]^2 dx] \; / \; [\int_R dx] .$$

We can partition this measure into two terms

$$J = V + B$$

where

$$V = \Omega\int_R V[\hat{y}(x)]dx, \quad B = \Omega\int_R [E\hat{y}_s(x)-\eta_t(x)]^2 dx$$

and

$$\Omega^{-1} = \int_R dx$$

It is easy to see the reason for the use of such a criterion by the relationship of $V$ and $B$ to estimation and model discrimination.

Notice that the primary aim of the type of investigation under consideration is to determine a relationship between the dependent variable $\eta$ and the independent variables $x_1,\ldots,x_p$ which reasonably represents the true relationship. It is the accuracy of the determined relationship and not any particular feature of the determined relationship that is important for many usages. Thus the problem is one of constructing a metric between the fitted function and the true function, defined, of course, over the region of interest.

As mentioned in the review of literature, a meaningful criterion considered by Box and Draper (1959) is the average of the squared deviations over the region R. This is the J criterion given above.

Once we agree upon a criterion of goodness of a design, it is desirable to obtain designs which are optimal and also allow the estimation of its value for each particular situation.

We will refer to designs having this property as min-average bias estimable designs. It is the aim of this work to determine such designs as well as to study their properties.

## B. Min-average Bias Estimable Designs

### 1. A basic result

Let the fitted model be a polynomial of degree $s$

$$\hat{y}_s(x) = \underline{x}_1'\underline{b}_1 \tag{3.4}$$

and the true underlying model a polynomial of degree $t > s$

$$\eta_t(x) = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2 \ , \tag{3.5}$$

where the least squares estimator of $\underline{\beta}_1$, $\underline{b}_1$, the design matrices $X_1$, $X = (X_1, X_2)$, and the vectors $\underline{\beta}_1$ and $\underline{\beta}_2$ all have the forms given in the previous section. Let the true model be the one expressed by 3.5. Then the value of the $J$ criterion for the fitted polynomial given by 3.4 is equal to

$$J = \Omega\sigma^2 \int_R x_1'(X_1'X_1)^{-1}x_1 \, dx + \Omega \int_R \underline{\beta}_2'(\underline{x}_2' - A\underline{x}_2')'(x_1'A - x_2')\underline{\beta}_2 \, dx,$$

where $A$ is as defined previously.

The bias component of $J$ can be written as

$$B = \underline{\beta}_2' \, \Delta \, \underline{\beta}_2 \ ,$$

with

$$\Delta = [W_3 - W_2'W_1^{-1}W_2] + [(X_1'X_1)^{-1}X_1'X_2 - W_1^{-1}W_2]'W_1[(X_1'X_1)^{-1}X_1'X_2 - W_1^{-1}W_2]$$

$$= \Delta_1 + \Delta_2 \ ,$$

where

$$W_1 = \Omega\int_R \underline{x}_1\underline{x}_1' dx, \quad W_2 = \Omega\int_R \underline{x}_1\underline{x}_2' dx, \quad W_3 = \Omega\int_R \underline{x}_2\underline{x}_2' \, dx$$

are the moments of a uniform distribution over the region R.

Box and Draper (1959) showed that $W_1$ and $\Delta_1$ are positive semi-definite, so the minimum B value is given by

$$\text{Min } B = \underline{\beta}_2'[W_3 - W_2'W_1^{-1}W_2]\underline{\beta}_2$$

and is achieved under the condition

$$(X_1'X_1)^{-1} X_1'X_2 = W_1^{-1}W_2 \, . \tag{3.6}$$

In particular, the minimum is achieved when

$$N^{-1}(X_1'X_1) = W_1 \quad \text{and} \quad N^{-1}(X_1'X_2) = W_2 \, . \tag{3.7}$$

Notice that $X_1'X_1$, $X_1'X_2$ and $X_2'X_2$ involve the cross-products of the design variables so $N^{-1}(X_1'X_1)$, $N^{-1}(X_2'X_2)$ are called "design moment matrices."

We can restate the condition given by 3.7 as follows. If a polynomial of degree s is fitted and the true polynomial has degree $t > s$, the minimum average square bias is achieved by designs with design moments up to order $s + t$ equal to the corresponding moments of a uniform or beta distribution Be(1,1) over the region of interest.

## 2. Characterization of the min-average bias estimable designs

The residual sum of squares of the model fitted according to 3.4 is given by

$$SS_R = y'(I-X_1'(X_1'X_1)^{-1}X_1)y \quad .$$

Since the true model is $\eta_t(x) = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2$

$$E(SS_R) = N\underline{\beta}_2'[X_2'X_2 - X_2'X_1(X_1'X_2)^{-1}X_1'X_2]\underline{\beta}_2 + \nu\sigma^2 \quad ,$$

where

$$\nu = N - s \quad .$$

Let us now impose on the design matrix for the complete model given by 3.5 the minimum bias conditions given by 3.6, and the additional condition $N^{-1}(X_2'X_2) = W_3$ so that we will have

$$N^{-1}(X_1'X_1) = W_1, \quad W^{-1}(X_1'X_2) = W_2, \text{ and } \quad N^{-1}(X_2'X_2) = W_3. \tag{3.8}$$

It will then follow that

$$N^{-1}(X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2) = W_3 - W_2'W_1^{-1}W_2$$

and

$$E\left(\frac{SS_R}{N}\right) = \underline{\beta}_2'[W_3 - W_2'W_1^{-1}W_2]\underline{\beta}_2 + \frac{\nu}{N}\sigma^2 \quad .$$

If we have an unbiased estimator $\hat{\sigma}^2$ of $\sigma^2$

$$E\left(\frac{SS_R}{N} - \frac{\nu}{N}\hat{\sigma}^2\right) = \underline{\beta}_2'[W_3 - W_2'W_1^{-1}W_2]\underline{\beta}_2 \quad ,$$

so the average square bias $B$ will be estimable.

We can now obtain an unbiased estimator for the average mean square error, or J-criterion, by adding to the expression given above an unbiased estimator of the average of the variance function of the fitted polynomial over the region R.

Let us summarize the result above as follows:

Theorem 3.1. Suppose we fit a polynomial of degree $s$ and the true underlying model is a polynomial of degree $t$, with $t > s$. If we have an unbiased estimator of $\sigma^2$, a sufficient condition to have the average square bias estimable, is that the design moments up to the order $2t$ be equal to the corresponding moments of a beta distribution $Be(1, 1)$ over the region of interest.

Under the condition above it follows that an unbiased estimator of the average square bias is given by

$$\hat{B}_{s,t} = \frac{SS_R}{N} - \frac{\nu}{N}\hat{\sigma}^2$$

where $SS_R$ is the residual sum of squares, $N$ is the total number of observations, $\nu = N - s$, and $\hat{\sigma}^2$ is an unbiased estimator of $\sigma^2$.

An unbiased estimator of the average mean square error is then

$$\hat{J} = \hat{V}_s + \hat{B}_{s,t}$$

where

$$\hat{V}_s = \Omega\hat{\sigma}^2 \int_R x_1'(X_1'X_1)^{-1}x_1 \, dx \ .$$

An additional property of the designs above is that they also minimize the average square bias.

An example of a case in which the use of min-average bias estimable designs leads to smaller average mean square error than by fitting the complete model is the following.

Let us assume that the true underlying model is a polynomial of degree $t$ and using a min-average bias estimable design we fit a polynomial of degree $s < t$. It follows from the above, that

$$\frac{NJ_s}{\Omega} = \sigma^2 + N\underline{\beta}_2'(W_3 - W_2'W_1^{-1}W_2)\underline{\beta}_2$$

and

$$\frac{NJ_t}{\Omega} = \sigma^2 t$$

Note that for the case under consideration

$$\frac{NJ_s}{\Omega} < \frac{NJ_t}{\Omega}$$

if and only if

$$\sigma^2 s + N\underline{\beta}_2'(W_3 - W_2'W_1^{-1}W_2)\underline{\beta}_2 < \sigma^2 t \ ,$$

i.e.,

$$N\underline{\beta}_2'(W_3 - W_2'W_1^{-1}W_2)\underline{\beta}_2 < \sigma^2(t-s) ,$$

and the SS for lack of it has expectation

$$\sigma^2(t-s) + N\underline{\beta}_2'(W_3 - W_2'W_1^{-1}W_2)\underline{\beta}_2 .$$

From the last two equations

$$\frac{NJ_s}{\Omega} < \frac{NJ_t}{\Omega}$$

if and only

$$F^* = \frac{E[\text{mean square for lack of fit}]}{E[\text{pure error mean square}]} < 2.$$

The above leads then to the following: if a min-average bias estimable design is used and if the ratio $F^*$ above is smaller than 2, one will have a smaller average mean square error by fitting the incomplete model. This suggests that one should use an actual F-ratio of 2 as a cut-off point in sequential fitting of a sequence of models.

To illustrate the results obtained in the previous section, consider the case of a single experimental variable.

## 3. Illustrative examples

We first take the simple case of fitting a linear relation

$$\eta_L(x) = \beta_0 + \beta_1 x$$

when the true relationship has the form

$$\eta_Q(x) = \beta_0 + \beta_1 x + \beta_2 x^2 \ .$$

If we have observations at points $x_1, x_2, \ldots, x_N$ of the interval $[-1, 1]$, the matrices involved are as follows:

$$X_1 = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} , \quad X_2 = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_N^2 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 \end{bmatrix} \ .$$

Hence we can write

$$X'X = N \begin{bmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{bmatrix} \ .$$

Therefore,

$$X_1'X_1 = N \begin{bmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{bmatrix} , \quad X_1'X_2 = N \begin{bmatrix} \mu_2 \\ \mu_3 \end{bmatrix} ,$$

$$(X_1'X_1)^{-1} = \frac{1}{N(\mu_2 - \mu_1^2)} \begin{bmatrix} \mu_2 & -\mu_1 \\ -\mu_1 & 1 \end{bmatrix}$$

$$(X_1'X_1)^{-1}X_1'X_2 = \frac{1}{(\mu_2-\mu_1^2)} \begin{bmatrix} \mu_2^2 - \mu_1\mu_3 \\ \mu_3 - \mu_1\mu_2 \end{bmatrix},$$

and

$$\frac{1}{N}[X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2]$$

$$= \mu_4 - \frac{1}{\mu_2-\mu_1^2}(\mu_2^3 + \mu_3^2 - 2\mu_1\mu_2\mu_3)$$

$$= \frac{1}{\mu_2-\mu_1^2}(\mu_4\mu_2 - \mu_4\mu_1^2 - \mu_2^3 - \mu_3^2 + 2\mu_1\mu_2\mu_3). \qquad (3.9)$$

Also we have

$$\begin{bmatrix} W_1' & W_2 \\ W_2 & W_3 \end{bmatrix} = \left[\begin{array}{cc:c} 1 & 0 & \nu_2 \\ 0 & \nu_2 & 0 \\ \hdashline \nu_2 & 0 & \nu_4 \end{array}\right]$$

where

$$\nu_2 = \frac{1}{2}\int_{-1}^{1} x^2\,dx = \frac{1}{3}, \quad \nu_4 = \frac{1}{2}\int_{-1}^{1} x^4\,dx = \frac{1}{5}.$$

We have therefore

$$W_1 = \begin{bmatrix} 1 & 0 \\ 0 & \nu_2 \end{bmatrix}, \quad W_1^{-1} = \frac{1}{\nu_2}\begin{bmatrix} \nu_2 & 0 \\ 0 & 1 \end{bmatrix},$$

$$W_2' = [\nu_2, 0], \quad W_3 = [\nu_4], \quad W_1^{-1}W_2 = \begin{bmatrix} \nu_2 \\ 0 \end{bmatrix},$$

so

$$W_3 - W_2'W_1^{-1}W_2 = \nu_4 - \nu_2^2 = \frac{4}{45} \quad .$$

Also

$$(X_1'X_1)^{-1} X_1'X_2 - W_1^{-1}W_2 =$$

$$= \frac{1}{\mu_2 - \mu_1^2} \begin{bmatrix} \mu_2^2 - \mu_1 \mu_3 \\ \mu_3 - \mu_1 \mu_2 \end{bmatrix} - \begin{bmatrix} \nu_2 \\ 0 \end{bmatrix}.$$

Hence

$$\Delta = \frac{4}{45} + \left[ \frac{1}{\mu_2 - \mu_1^2} \begin{bmatrix} \mu_2^2 & \mu_1\mu_3 \\ \mu_3 - \mu_1\mu_2 \end{bmatrix} - \begin{bmatrix} \nu_2 \\ 0 \end{bmatrix} \right]' \begin{bmatrix} 1 & 0 \\ 0 & \nu_2 \end{bmatrix}$$

$$\left[ \frac{1}{\mu_2 - \mu_1^2} \begin{bmatrix} \mu_2^2 - \mu_1\mu_3 \\ \mu_3 - \mu_1\mu_2 \end{bmatrix} - \begin{bmatrix} \nu_2 \\ 0 \end{bmatrix} \right],$$

or

$$\Delta = \frac{4}{45} + \left(\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2} - \nu_2\right)^2 + \nu_2 \left(\frac{\mu_2 - \mu_1\mu_2}{\mu_2 - \mu_1^2}\right)^2 .$$

Substituting for $\nu_2$ yields

$$B_{1,2} = \left[\frac{4}{45} + \left(\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2} - \frac{1}{3}\right)^2 + \frac{1}{3}\left(\frac{\mu_3 - \mu_1\mu_2}{\mu_2 - \mu_1^2}\right)^2\right] \beta_2^2 . \qquad (3.10)$$

From the design moment conditions given by 3.8

$$\mu_1 = 0, \quad \mu_2 = \frac{1}{3}, \quad \mu_3 = 0, \quad \mu_4 = \frac{1}{5}, \tag{3.11}$$

and it follows that the right-hand sides of 3.9 and 3.10 are equal, and hence $B_{1,2}$ is estimable if we have an independent unbiased estimator of $\sigma^2$.

Also note that from 3.10 it follows that the expression of the average square bias for a symmetrical design is given by

$$B_{1,2} = [\frac{4}{45} + (\mu_2 - \frac{1}{3})^2]\beta_2^2 \ . \tag{3.12}$$

We now consider the case where we fit a quadratic polynomial and the true underlying model is a cubic

$$\eta_C(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \ .$$

We consider only symmetrical designs. The X'X matrix for the complete model can be partitioned as

$$\left[\begin{array}{cc:c} 1 & 0 & \mu_2 & 0 \\ 0 & \mu_2 & 0 & \mu_4 \\ \mu_2 & 0 & \mu_4 & 0 \\ \hdashline 0 & \mu_4 & 0 & \mu_6 \end{array}\right]$$

where the upper left submatrix is the $X_1'X_1$ matrix corresponding to the fitted quadratic model.

$$(X_1'X_1)^{-1} X_1'X_2 = [0, \mu_4/\mu_2, 0]'$$

and

$$N^{-1}(X_2'X_2 - X_2'X_1(X_1'X_1)^{-1} X_1'X_2) = \mu_6 - \frac{\mu_4^2}{\mu_2} . \qquad (3.13)$$

Also we have

$$\begin{bmatrix} W_1 & W_2 \\ W_2' & W_3 \end{bmatrix} = \left[\begin{array}{ccc:c} 1 & 0 & \nu_2 & 0 \\ 0 & \nu_2 & 0 & \nu_4 \\ \nu_2 & 0 & \nu_4 & 0 \\ \hdashline 0 & \nu_4 & 0 & \nu_6 \end{array}\right]$$

where as above, $\nu_2$, $\nu_4$ and $\nu_6$ stand for the moments of the beta distribution Be(1,1) over the [-1, 1] interval, or:

$$\nu_2 = \frac{1}{3} , \quad \nu_4 = \frac{1}{5} \quad \text{and} \quad \nu_6 = \frac{1}{7} .$$

Note that

$$W_1 = \begin{bmatrix} 1 & 0 & \nu_2 \\ 0 & \nu_2 & 0 \\ \nu_2 & 0 & \nu_4 \end{bmatrix}, \quad W_1^{-1} = \begin{bmatrix} \frac{\nu_4}{\nu_4-\nu_2^2} & 0 & \frac{-\nu_2}{\nu_4-\nu_2^2} \\ 0 & \frac{1}{\nu_2} & 0 \\ \frac{-\nu_2}{\nu_4-\nu_2^2} & 0 & \frac{1}{\nu_4-\nu_2^2} \end{bmatrix},$$

$$W_2' = [0,\ \nu_4,\ 0],\quad W_3 = [\nu_6]\ .$$

Hence

$$W_1^{-1}W_2 = [0,\ \nu_4/\nu_2,\ 0]'\ ,$$

so

$$W_3 - W_2'W_1^{-1}W_2 = \nu_6 - \frac{\nu_4^2}{\nu_2} = \frac{4}{175}\ .$$

Also

$$(X'_1X_1)^{-1}X_1'X_2 - W_1^{-1}W_2 = [0,\ \mu_4/\mu_2 - \nu_4/\nu_2,\ 0]'\ .$$

Hence

$$\Delta = \frac{4}{175} + \frac{1}{\nu_2}\left(\frac{\mu_4}{\mu_2} - \frac{\nu_4}{\nu_2}\right)^2\ .$$

Substituting for $\nu_4$ and $\nu_2$ yields

$$B_{2,3} = \left[\frac{4}{175} + \frac{1}{3}\left(\frac{\mu_4}{\mu_2} - \frac{3}{5}\right)^2\right]\beta_3^2\ , \tag{3.14}$$

the general expression of the average mean square bias for a symmetrical design when we fit a quadratic polynomial and the true model is a cubic.

Notice that by making the design moments $\mu_2$, $\mu_4$, $\mu_6$ equal to the corresponding moments of a beta distribution Be(1,1) over the [-1, 1] interval, we will have the expression given by 3.13 equal to the one given by 3.14 so that $B_{2,3}$ is estimable assuming we have an unbiased estimator of $\sigma^2$.

Finally, consider fitting a linear polynomial when the

true one is cubic. We consider only symmetrical designs, and find the relevant matrices to be as follows:

$$X'X = N \begin{bmatrix} 1 & 0 & \mu_2 & 0 \\ 0 & \mu_2 & 0 & \mu_4 \\ \mu_2 & 0 & \mu_4 & 0 \\ 0 & \mu_4 & 0 & \mu_6 \end{bmatrix}.$$

$$X_1'X_1 = N \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix}; \quad (X_1X_1)^{-1} = \frac{1}{N\mu_2} \begin{bmatrix} \mu_2 & 0 \\ 0 & 1 \end{bmatrix},$$

$$X_1'X_2 = N \begin{bmatrix} \mu_2 & 0 \\ 0 & \mu_2 \end{bmatrix},$$

$$(X_1'X_1)^{-1} X_1'X_2 = \frac{1}{\mu_2} \begin{bmatrix} \mu_2^2 & 0 \\ 0 & \mu_4 \end{bmatrix}$$

and

$$X_1'X_2 = \begin{bmatrix} \mu_4 & 0 \\ 0 & \mu_6 \end{bmatrix}.$$

Hence

$$\frac{1}{N}[X_2'X_2 - X_2'X_1(X_1'X_1)^{-1} X_1'X_2] =$$

$$= \left[ \begin{bmatrix} \mu_4 & 0 \\ 0 & \mu_6 \end{bmatrix} - \begin{bmatrix} \mu_2 & 0 \\ 0 & \mu_4 \end{bmatrix} \frac{1}{\mu_2 - \mu_1^2} \begin{bmatrix} \mu_2^2 & 0 \\ 0 & \mu_4 \end{bmatrix} \begin{bmatrix} \mu_2 & 0 \\ 0 & \mu_4 \end{bmatrix} \right].$$

Also

$$W_1^{-1}W_2 = \begin{bmatrix} 1 & 0 \\ 0 & \nu_2 \end{bmatrix}^{-1} \begin{bmatrix} \nu_2 & 0 \\ 0 & \nu_4 \end{bmatrix} = \frac{1}{\nu_2} \begin{bmatrix} \nu_2^2 & 0 \\ 0 & \nu_4 \end{bmatrix}$$

$$\Delta_1 = W_3 - W_2'W_1^{-1}W_2 = \frac{1}{\nu_2} \begin{bmatrix} \nu_2\nu_4 - \nu_2^3 & 0 \\ 0 & \nu_2\nu_6 - \nu_4^2 \end{bmatrix}.$$

Hence

$$\underline{\beta}_2\Delta_1\underline{\beta}_2 = (\nu_4 - \nu_2^2)\beta_2^2 + (\nu_6 - \frac{\nu_4^2}{\nu_2})\beta_3^2 .$$

It follows that with $\nu_2 = \frac{1}{3}$ and $\nu_4 = \frac{1}{5}$,

$$\underline{\beta}_2\Delta_1\underline{\beta}_2 = \frac{4}{45}\beta_2^2 + \frac{4}{175}\beta_3^2 . \qquad (3.15)$$

Also note that

$$\Delta_2 = \{[(X_1'X_1)^{-1}X_1'X_2 - W_1^{-1}W_2]'W_1[(X_1'X_1)^{-1}X_1'X_2 - W_1^{-1}W_2]\} =$$

$$= \left[ \frac{1}{\mu_2} \begin{bmatrix} \mu_2^2 & 0 \\ 0 & \mu_4 \end{bmatrix} - \frac{1}{\nu_2} \begin{bmatrix} \nu_2^2 & 0 \\ 0 & \nu_4 \end{bmatrix} \right]' \begin{bmatrix} 1 & 0 \\ 0 & \nu_2 \end{bmatrix} \left[ \frac{1}{\mu_2} \begin{bmatrix} \mu_2^2 & 0 \\ 0 & \mu_4 \end{bmatrix} - \frac{1}{\nu_2} \begin{bmatrix} \nu_2^2 & 0 \\ 0 & \nu_4 \end{bmatrix} \right]$$

so that for $\nu_2$ and $\nu_4$ as above

$$\Delta_2 = \begin{bmatrix} (\mu_2 - \frac{1}{3})^2 & 0 \\ 0 & \frac{1}{3}(\frac{\mu_4}{\mu_2} - \frac{3}{5})^2 \end{bmatrix} .$$

If we now impose the moment conditions given by 3.8 or, $\mu_2 = \frac{1}{3}$ and $\mu_4 = \frac{1}{5}$, it will then follow that $\Delta_2 = 0$ and

$$\frac{1}{N}(X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2) = \Delta_1 .$$

Thus, B is estimable if we also have an unbiased estimator of $\sigma^2$.

Note that from the above equation the general expression of $B^2_{1,3}$ for symmetrical designs is

$$B_{1,3} = [\frac{4}{45} + (\mu_2 - \frac{1}{3})^2]\beta_2^2 + [\frac{4}{175} + \frac{1}{3}(\frac{\mu_4}{\mu_2} - \frac{3}{5})^2]\beta_3^2 \qquad (3.16)$$

hence for symmetrical designs we have

$$B^2_{1,3} = B^2_{1,2} + B^2_{2,3} .$$

That is, the average mean square bias resulting from fitting a linear polynomial when the true model is cubic, is equal to the sum of the corresponding expressions for fitting a linear polynomial when the true model is a quadratic polynomial and for fitting a quadratic polynomial when the true model is a cubic. The above property does not hold in general.

## 4. Some min-average bias estimable designs

We will now obtain some one-dimensional min-average bias estimable designs.

Consideration will be given to designs for fitting polynomials of up to the third degree because those are the cases of more common practical use, although these methods can be applied to polynomials of any degree. Without loss of generality, as we have been doing earlier, we will consider only symmetrical designs in the $[-1, 1]$ interval of the real line.

Suppose we are going to fit a first degree polynomial but are concerned with appropriateness of this model compared to a second degree model.

To have the minimum average square bias estimable, we impose the moments conditions given by 3.8 which can be written as

$$\mu_1 = N^{-1} \sum_{i=1}^{I} \eta_i x_i = 0, \quad \mu_2 = N^{-1} \sum_{i=1}^{I} \eta_i x_i^2 = \frac{1}{3},$$

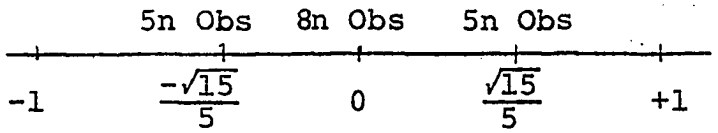
$$\mu_3 = N^{-1} \sum_{i=1}^{I} \eta_i x_i^3 = 0, \quad \mu_4 = N^{-1} \sum_{i=1}^{I} \eta_i x_i^4 = \frac{1}{5},$$

where $x_i$ are one dimensional x-allocations, $n_i$ is the number of observations taken at $x_i$ and $I$ is the number of different allocations.

We consider one-dimensional symmetrical designs with 3, 4 or 5 allocation points, although of course we do not

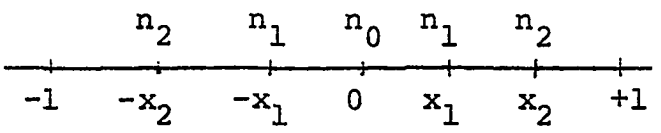
need to restrict ourselves only to such situations.

It is easy to prove that the only min-average bias estimable designs with 3 allocation points are the ones of the form




where $N = 18n$, $n = 1,2,\ldots,$ is the total number of observations.

Symmetrical designs with 4 or 5 different allocations are of the form




.

They can be obtained by solving systems such as

$$6n_1x_1^2 + 6n_2x_2^2 = N$$

$$10n_1x_1^4 + 10n_2x_2^4 = N$$

$$2(n_1+n_2) = N$$

where $N$ and $n_1$(or $n_2$) are fixed in advance, and $n_2$(or $n_1$), $x_1$ and $x_2$ are meaningful solutions.

Table 3.1 contains a list of some designs of interest.

We now consider the case where we take the fitted curve

Table 3.1. Min-average bias estimable designs for linear fits

| $x_1$ | $x_2$ | $n_1$ | $n_2$ | $n_0$ | N |
|---|---|---|---|---|---|
| $\frac{5+\sqrt{20}}{15}$ | $\frac{5-\sqrt{20}}{15}$ | n | n | 0 | 4n |
| $\frac{5+\sqrt{11}}{20}$ | $\frac{5-\sqrt{11}}{20}$ | n | n | n | 5n |
| $\frac{5+\sqrt{5}}{10}$ | $\frac{5-\sqrt{5}}{10}$ | n | n | 2n | 6n |
| $\frac{5+2\sqrt{10}}{15}$ | $\frac{5-2\sqrt{10}}{15}$ | n | 2n | 0 | 6n |
| $\frac{35+\sqrt{35}}{60}$ | $\frac{35-\sqrt{35}}{60}$ | n | n | 3n | 7n |
| $\frac{10+\sqrt{10}}{20}$ | $\frac{10-\sqrt{10}}{20}$ | n | 2n | 3n | 9n |
| $\frac{5+3\sqrt{15}}{40}$ | $\frac{5-3\sqrt{15}}{40}$ | 2n | 2n | n | 9n |

to be a second degree polynomial but the true response curve is a third degree polynomial. The design moments conditions in this case are

$$\mu_1 = N^{-1} \sum_{i=1}^{I} n_i x_i = 0, \quad \mu_2 = N^{-1} \sum_{i=1}^{I} n_i x_i^2 = \frac{1}{3},$$

$$\mu_3 = N^{-1} \sum_{i=1}^{I} n_i x_i^3 = 0, \quad \mu_4 = N^{-1} \sum_{i=1}^{I} n_i x_i^4 = \frac{1}{5},$$

$$\mu_5 = N^{-1} \sum_{i=1}^{I} n_i x_i^5 = 0, \quad \mu_6 = N^{-1} \sum_{i=1}^{I} n_i x_i^6 = \frac{1}{7}.$$

Those conditions include the ones for first versus second degree polynomials, so, as mentioned earlier, these conditions are sufficient to maintain estimability and minimum bias in the sequential discrimination of models from the first to the third degree.

For designs with symmetrical allocation at six or seven points our system of equations is:

$$2n_1x_1^2 + 2n_2x_2^2 + 2n_3x_3^2 = \frac{N}{3}$$

$$2n_1x_1^4 + 2n_2x_2^4 + 2n_3x_3^4 = \frac{N}{5}$$

$$2n_1x_1^6 + 2n_2x_2^6 + 2n_3x_3^6 = \frac{N}{7}$$

$$2(n_1+n_2+n_3) + n_0 = N .$$

If we restrict ourselves to designs with an equal number

of observations n at the symmetric points, our system will be:

$$x_1^2 + x_2^2 + x_3^2 = \frac{N}{6}$$

$$x_1^4 + x_2^4 + x_3^4 = \frac{N}{10}$$

$$x_1^6 + x_2^6 + x_3^6 = \frac{N}{14}$$

$$6n + n_0 = N,$$

where as above $n_0$, is the number of observations at $x = 0$.

We can reduce the above systems to the general form:

$$t_1 + t_2 + t_3 = c_1$$

$$t_1^2 + t_2^2 + t_3^2 = c_2$$

$$t_1^3 + t_2^3 + t_3^3 = c_3$$

with

$$t_1, t_2, t_3 \geq 0.$$

The general approach to solving these systems is as follows. Write

$$(t_1+t_2+t_3)^3 = -2(t_1^3+t_2^3+t_3^3) + 3(t_1^2+t_2^2+t_3^2)(t_1+t_2+t_3) - 5t_1t_2t_3,$$

or

$$c_1^3 = -2c_3 + 3c_2c_1 - 5t_1t_2t_3$$

from which we get

$$t_1t_2t_3 = (c_1^3 + 2c_3 - 3c_3c_1)/6 ,$$

and similarly from

$$(t_1 + t_2 + t_3)^2$$

$$t_1t_2 + t_1t_3 + t_2t_3 = (c_1^2 - c_2)/2 .$$

Now let us consider $t_1$, $t_2$ and $t_3$ as solutions of a third degree equation,

$$t^3 + a_1t^2 + a_2t + a_3 = 0 .$$

The well-known relationship between coefficients and roots of an 3-rd degree polynomial give

$$t_1 + t_2 + t_3 = -a_1$$

$$t_1t_2 + t_1t_3 + t_2t_3 = a_2$$

$$t_1t_2t_3 = -a_3 .$$

So that

$$a_1 = -c_1$$

$$a_2 = (c_1^2-c_2)/2$$

$$a_3 = -(c_1^3+2c_3-3c_2c_1)/6$$

and $t_1$, $t_2$ and $t_3$ are determined by solving the above equation.

We now list some designs obtained as above.

Table 3.2. Min-average bias estimable designs for quadratic fits

| $x_1$ | $x_2$ | $x_3$ | $n_1$ | $n_2$ | $n_3$ | $n_0$ | N |
|---|---|---|---|---|---|---|---|
| .36 | .42 | .86 | n | n | n | n | 7n |
| .32 | .52 | .88 | n | n | n | 2n | 8n |
| .44 | .57 | .89 | n | n | n | 3n | 9n |

We cannot increase the number of observations at zero to obtain designs with a greater number of observations because the corresponding systems will have complex solutions. We may consider, however, designs with more numbers of allocation points and apply the above general method.

## C. Minimum Average Weighted Bias Designs

### 1. General results

We will generalize the work on the minimization of the mean average square bias done by Box and Draper (1959), by considering the average mean square bias over a pre-chosen region of interest weighted with a non-negative function, satisfying certain integrability conditions, as the criterion of optimality. We will restrict ourselves to the one-dimensional case, although as we will see in the following the results of this section can be easily generalized to $k$ dimensions.

The use of a weight function has been suggested by G. E. P. Box.

Suppose we take observations at values of $x$ equal to $x_1, x_2, \ldots, x_n$, and that we choose for fitting an s-th degree polynomial

$$\hat{y}_s(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \ldots + \hat{\beta}_s x^s$$

or

$$\hat{y}_s(x) = \underline{x}_1' \hat{\underline{\beta}}_1$$

with

$$\underline{x}_1' = (1, x, \ldots, x^s), \quad \text{and} \quad \underline{\beta}_1 = (\beta_0, \beta_1, \ldots, \beta_s).$$

Let the true underlying polynomial model be

$$\eta_t(x) = \beta_0 + \beta_1 x + \ldots + \beta_t x^t$$

or

$$\eta_t(x) = \underline{x}_1' \underline{\beta}_1 + \underline{x}_2' \underline{\beta}_2$$

with

$$\underline{x}_2' = (x^{s+1}, \ldots, x^t), \quad \text{and} \quad \underline{\beta}_2' = (\beta_{s+1}, \ldots, \beta_t).$$

The design matrix corresponding to the complete model can be partitioned as

$$X = (X_1, X_2)$$

where

$$X_1 = \begin{bmatrix} 1 & x_1 & \cdots & x_1^s \\ \vdots & & & \vdots \\ 1 & x_N & \cdots & x_N^s \end{bmatrix},$$

and

$$X_2 = \begin{bmatrix} x_1^{s+1} & \cdots & x_1^t \\ \vdots & & \vdots \\ x_N^{s+1} & \cdots & x_N^t \end{bmatrix}.$$

Hence the fitted value for the true yield at the point $x$ is

$$\hat{y}_s(x) = \underline{x}_1'[\underline{\beta}_1+(X_1'X_1)^{-1}X_1'X_2\underline{\beta}_2+(X_1'X_1)^{-1}X_1'\underline{e}] ,$$

the error in this fitted value is given by

$$\hat{y}_s(x)-\eta_t(x) = [\underline{x}_1'(X_1'X_1)^{-1}X_1'X_2-\underline{x}_2']\beta_2+x_1'(X_1'X_1)^{-1}X_1'\underline{e} ,$$

and the bias is

$$E\hat{y}_s(x)-\eta_t(x) = [x_1'(X_1'X_1)^{-1}X_1'X_2-x_2']\underline{\beta}_2 .$$

It follows immediately that the $(\text{bias})^2$ function is

$$B^2_{s,t}(x) = \underline{\beta}_2'[\underline{x}_1'A-\underline{x}_2']'[\underline{x}_1'A-\underline{x}_2']\underline{\beta}_2 \qquad (3.17)$$

where

$$A = (X_1'X_1)^{-1}X_1'X_2 .$$

We define the mean average weighted square bias as

$$B^{w}_{s,t} = \int_a^b [E\hat{y}_s(x) - \eta_t(x)]^2 w(x)dx \ ,$$

where $w(x) \geq 0$ over $[a,b]$ and is such that the moments with respect to it

$$\mu_n = \int_a^b x^n w(x)dx, \quad n = 0,1,2,... \tag{3.18}$$

exist and are finite.

From 3.17 it follows that we can express $B^{w}_{s,t}$ as,

$$B^{w}_{s,t} = \int_a^b \underline{\beta}_2' [\underline{x}_1'A - \underline{x}_2'] ' [\underline{x}_1'A - \underline{x}_2'] \underline{\beta}_2 w(x)dx_1$$

or

$$B^{w} = \underline{\beta}_2' \Delta^{w} \underline{\beta}_2$$

where

$$\Delta^{w} = (A'W_1A - W_2'A - A'W_2 + W_3)$$

with

$$W_1 = \int_a^b \underline{x}_1\underline{x}_1' w(x)dx, \quad W_2 = \int_a^b \underline{x}_1\underline{x}_2' w(x)dx,$$

and

$$W_3 = \int_a^b \underline{x}_2\underline{x}_2' w(x)dx \ .$$

We can also express $\Delta^{w}$ as

$$\Delta^{w} = (W_3 - W_2'W_1^{-1}W_2) + (A - W_1^{-1}W_2)'W_1(A - W_1^{-1}W_2)$$

or

$$\Delta^{w} = \Delta_1^{w} + \Delta_2^{w} \ ,$$

with $\Delta_1^{w}$ independent of the design. Note now that

$$\int_a^b [(\underline{x}_1', \underline{x}_2')\underline{\beta}]^2 w(x)\,dx = \underline{\beta}' \int_a^b \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix} [\underline{x}_1', \underline{x}_2'] w(x)\,dx\ \underline{\beta}$$

$$= \underline{\beta}' \begin{bmatrix} W_1 & W_2 \\ W_2' & W_3 \end{bmatrix} \underline{\beta} .$$

Hence

$$W_1 \quad \text{and} \quad \begin{bmatrix} W_1 & W_2 \\ W_2' & W_3 \end{bmatrix}$$

are both positive semi-definite. Since we can write

$$\begin{bmatrix} W_1 & 0 \\ 0 & \Delta_1^W \end{bmatrix} = \begin{bmatrix} W_1 & 0 \\ 0 & W_3 - W_2' W_1^{-1} W_2 \end{bmatrix}$$

$$= U' \begin{bmatrix} W_1 & W_2 \\ W_2' & W_3 \end{bmatrix} U$$

where

$$U = \begin{bmatrix} I & -W_1^{-1} W_2 \\ 0 & W_3 \end{bmatrix} ,$$

it follows that $\Delta_1^W$ is also positive semi-definite.

Thus, $\Delta_1^W$ as well as $W_1$ are positive semi-definite. It then follows that we will minimize $\underline{\beta}_2' \Delta^2 \underline{\beta}_2$ independently of the value of $\underline{\beta}_2$ by taking

$$A = W_1^{-1} W_2$$

or

$$(X_1'X_1)^{-1}X_1'X_2 = W_1^{-1}W_2 \ .$$

In particular

$$N^{-1}(X_1'X_1) = W_1, \quad \text{and} \quad N^{-1}(X_1'X_2) = W_2$$

satisfy this condition.

The proof above closely follows the one given by Box and Draper (1959) for the minimization of the average mean square bias. Notice that its generalization to the multi-dimensional case is immediate.

The generalization to the multi-dimensional case is obvious in that the weight function is to be defined over the chosen multi-dimensional region, and the averaging is over this region. However, the implementation of the prescription is not obvious, because while there exists an extensive theory of polynomials of a single variable over a segment of the real line, there appears to be little theory of polynomials over multi-dimensional regions that are not generalized rectangles or hyperspheres.

We can summarize the above as follows:

Theorem 3.2. Suppose a polynomial of degree $s$ is fitted and the true underlying functional relationship is a polynomial of degree $t$, with $t > s$. A necessary and

sufficient condition to minimize the $B^W$ criterion defined above is that

$$(X_1'X_1)^{-1}X_1'X_2 = W_1^{-1}W_2$$

where $W_1$ and $W_2$ are obtained substituting in the design moment matrices $N^{-1}(X_1'X_1)$ and $N^{-1}(X_1'X_2)$ the corresponding moments over the chosen region with respect to $w(x)$. A particular solution for the equation above is

$$N^{-1}(X_1'X_1) = W_1$$

and

$$N^{-1}(X_1'X_2) = W_2 \ .$$

This is equivalent to making the design moments up to the order $s+t$ equal to the corresponding moments with respect to $w(x)$.

A special case of the theorem above which deserves consideration for its relationship to orthogonal polynomials is the situation where the fitted polynomial has degree $s$ and the true underlying polynomial model has degree $s+1$. If the $s$-th degree polynomial is fitted taking observations at only $s+1$ allocation points, the expression of the $(\text{bias})^2$ function is

$$B^2_{s,s+1}(x) = \pi^2_{s+1}(x)\beta^2_{s+1} \tag{3.19}$$

where

$$\pi_{s+1}(x) = (x-x_1)\ldots(x-x_{s+1}).$$

Hence, the average weighted square bias can be expressed as

$$B^W_{s,s+1} = \beta^2_{s+1}\int_a^b \pi^2_{s+1}(x)w(x)dx \tag{3.20}$$

where $w(x)$ satisfy the conditions mentioned previously.

It follows that the polynomial $\pi_{s+1}(x)$ which minimizes the integral above is the polynomial of degree $s+1$ and leading coefficient 1 which is orthogonal to $w(x)$. From the above we can state the following result.

Theorem 3.3. Suppose a polynomial of degree $s$ is fitted and the true underlying model is a polynomial of degree $s+1$. The design which places the observations at the zeros of a $(s+1)$-th degree polynomial orthogonal with respect to the weight function $w(x)$ is optimal with respect to $B^W_{s,s+1}$.

A consequence of the above result is of special interest, because it relates probability density functions to orthogonal polynomials. It is easy to see from Theorem 3.3 and properties of orthogonal polynomials that the following result holds.

Theorem 3.4. Let $w(x)$ be one of the following weight functions:

(i) density of a beta $B(\alpha,\beta)$ distribution and $[a,b] = [-1, 1]$, $\alpha>0$, $\beta>0$;

(ii) density of a gamma distribution $G(\alpha)$ and $[a,b] = [0, \infty)$, $\alpha>0$;

(iii) density of a normal $N(0, \frac{1}{2})$ distribution and $[a,b] = (-\infty,+\infty)$.

Then the average weighted square bias $B^{w}_{s,s+1}$ is minimized by the design which places the observations at the zeros of the polynomials:

(i) $P^{(\alpha,\beta)}_{s+1}(x)$, the $(s+1)$-th Jacobi polynomial;

(ii) $L^{(\alpha)}_{s+1}(x)$, the $(s+1)$-th Laguerre polynomial;

and (iii) the Hermite polynomial $H_{s+1}(x)$ .

A listing of the zeros of the various orthogonal polynomials considered above is given in Abramovitz and Stegun (1968).

## 2. Additional properties of the min-average bias designs

Note that putting $\alpha=\beta=1$ in part (i) of Theorem 3.4, we will have the mean average bias criterion considered in the previous sections of this chapter and the optimal allocation given by this theorem is a design which places the

observations at the zeros of the Legendre polynomial of degree s+1. It follows from the above that the polynomial with leading coefficient 1 in the expression 3.19 for $B_{s,s+1}$ is given by

$$\pi_{s+1}(x) = \frac{2^{s+1}[(s+1)!]^2}{(2s+2)!} P_{s+1}(x) ,$$

where $P_{s+1}(x)$ is the Legendre polynomial of degree s+1. Since (Sansone, 1959),

$$|P_{s+1}(x)| < 1 \quad \text{for} \quad |x| < 1, \quad P^2_{s+1}(\pm 1) = 1$$

and

$$\int_{-1}^{1} P^2_{s+1}(x)\, dx = \frac{2}{2s+3} ,$$

it follows that

$$\max_{-1 \le x \le 1} \pi^2_{s+1}(x) = \left[\frac{2^{s+1}[(s+1)!]^2}{(2s+2)!}\right]^2$$

is attained at the points $\pm 1$, and

$$\frac{1}{2}\int_{-1}^{1} \pi^2_{s+1}(x)\, dx = \frac{1}{2s+3}\left[\frac{2^{s+1}[(s+1)!]^2}{(2s+2)!}\right]^2 .$$

It follows from the above, 3.19 and 3.20 that the following result holds.

Theorem 3.5. Suppose a polynomial of degree s is fitted and the true underlying polynomial model has degree s+1. Then the design which places the observations at the zeros of the Legendre polynomial of degree s+1 is optimal with

respect to the average square bias criterion. The expression for the $(\text{bias})^2$ function and the mean average square bias are given respectively by

$$B_{s,s+1}(x) = \left[\frac{2^{s+1}(s+1)!^2}{(2s+2)!}\right]^2 P_{s+1}^2(x),$$

and

$$B_{s,s+1} = \frac{1}{2s+3}\left[\frac{2^{s+1}[(s+1)!}{(2s+2)!}\right]^2]\beta_{s+1}^2 .$$

The maximum of the $(\text{bias})^2$ function over the interval $[-1, 1]$ is attained at the points $-1$ and $1$, and is given by

$$\max_{-1\leq x\leq 1} B_{s,s+1}^2(x) = \left[\frac{2^{s+1}\,[(s+1)!]^2}{(2s+2)!}\right]^2 \beta_{s+1}^2 .$$

As we will see in the following example, in general, the above designs will not be optimal if a model of degree $d < s$ is fitted and the true model has degree $d + 1$. Also, they do not necessarily satisfy the condition for the estimation of the minimum average square bias.

In the illustrative example given at Section A.3 of this chapter the fitted model is a quadratic polynomial but the true model is a cubic polynomial.

The general conditions for minimum average bias, imply that for the case under consideration

$$\frac{\mu_4}{\mu_2} = \frac{\nu_4}{\nu_2}$$

where $\nu_2 = 1/3$ and $\nu_4 = 1/5$ are the moments of the beta distribution Be(1, 1), or

$$\frac{\mu_4}{\mu_2} = \frac{3}{5} . \tag{3.21}$$

As mentioned earlier, a particular solution for the above is

$$\mu_2 = \nu_2 \quad \text{and} \quad \mu_4 = \nu_4 . \tag{3.22}$$

Let us now consider the Legendre allocation for the quadratic fit. The zeros of the Legendre polynomial of degree 3 are

$$x_1 = 0, \quad x_2 = \frac{+\sqrt{15}}{5} , \quad \text{and} \quad x_3 = \frac{-\sqrt{15}}{5}$$

and if we take an equal number of observations at each of the above points, the corresponding design moments are

$$\mu_1 = 0, \ \mu_2 = \frac{2}{5}, \ \mu_3 = 0, \ \mu_4 = \frac{6}{25} , \text{ and } \ \mu_6 = \frac{18}{125} .$$

It is easy to see that the design moments above satisfy the general conditions given by 3.21 but do not satisfy the $\mu_2 = 1/3$ condition given by 3.11. Hence these designs will not minimize the average square bias if the fitted model is linear and the true model is quadratic. Note also that the design above does not satisfy the condition for the estimation of the minimum average bias, i.e., $\mu_6 \neq \nu_6 = 1/7$.

## 3. Relationship between the minimax bias designs and a special average weighted square bias design

It is easy to see from Theorem 3.2, that an optimal allocation with respect to $B^{W}_{s,s+1}$ is not necessarily optimal if the fitted model has degree $d < s$ and the true model has degree $d+1$. This was just exemplified for allocation at the zeros of a Legendre polynomial and also holds for all the other orthogonal polynomials except the Chebyshev, which from Theorem 3.4 (i) is optimal for the case where the weight function in $B^{W}_{s,s+1}$ is the beta $Be(1/2, 1/2)$ density function. The above follows from the fact that the Chebyshev quadrature formula for an equal allocation at the zeros of $T_{s+1}(x)$ has design moments up to the order $2s+1$ equal to the corresponding moments of the beta $Be(1/2, 1/2)$ distribution defined over the interval under consideration. From Theorem 3.2, this property also leads to the result that the Chebyshev allocation at the zeros of $T_n(x)$ with $n \geq [(s+t+1)/2]$ is optimal with respect to the $B^{W}$ where $w(x)$ is the beta $Be(1/2, 1/2)$ density function. As proved in the second chapter, the Chebyshev allocation is also optimal for the maximum bias criterion.

The properties of the Chebyshev allocation mentioned above, as well as the fact that the min-average bias designs obtained by Box and Draper (1959), have variance and bias functions which assume large values at the end portions of the interval, motivates a preference for the beta $Be(1/2, 1/2)$

density function in the $B^W$ criterion over the ones which give more weight at the extreme portions of the interval, and suggests a more detailed study of the properties of the corresponding optimal designs. From the remarks above and Theorem 3.4 the following result holds.

Theorem 3.6. Let us consider as criterion of optimality the mean average square bias weighted as follows,

$$B^W_{s,t} = \frac{1}{\pi} \int_{-1}^{1} \frac{[E\hat{y}_s(x) - \eta_t(x)]^2}{\sqrt{1-x^2}} \, dx \ .$$

A necessary and sufficient condition to minimize $B^W_{s,t}$ is that

$$(X_1'X_1)^{-1}X_1'X_2 = W_1^{-1}W_2$$

where $W_1$ and $W_2$ are obtained by substituting in the design moments matrices $N^{-1}(X_1'X_1)$ and $N^{-1}(X_1'X_2)$ the corresponding moments of the beta distribution $Be(1/2, 1/2)$. A particular solution for the equation above is

$$N^{-1}(X_1'X_1) = W_1 \quad \text{and} \quad N^{-1}(X_1'X_2) = W_2 \ .$$

This is equivalent to making the design moments up to the order $s+t$ equal to the corresponding moments of the beta $Be(1/2, 1/2)$ distribution under consideration.

We now notice that the procedure for obtaining necessary and sufficient conditions to minimize the maximum of $B^2_{s,s+1}(x)$

given in Section B.3 of the second chapter is equivalent to obtaining designs which satisfy the following matrix equation

$$(X_1'X_1)^{-1}X_1'X_2 = W_1^{-1}W_2$$

where

$$N^{-1}(X_1'X_1) = \begin{bmatrix} 1 & \mu_1 & \cdots & \mu_s \\ \vdots & & & \vdots \\ \mu_s & \mu_{s+1} & \cdots & \mu_{2s} \end{bmatrix}, \quad N^{-1}(X_1'X_2) = \begin{bmatrix} \mu_{s+1} \\ \vdots \\ \mu_{2s+1} \end{bmatrix},$$

and $W_1$ and $W_2$ are obtained by substituting the design moments in the $N^{-1}(X_1'X_1)$ and $N^{-1}(X_1'X_2)$ matrices by the corresponding moments of the beta $Be(1/2, 1/2)$ distribution defined over the interval of interest.

On comparing the last conditions with the ones given in Theorem 3.6, it is immediate that the designs which minimize $B^w_{s,s+1}$ with $w(x) = \frac{1}{\pi}(1-x^2)^{-1/2}$ and the ones which minimize the maximum of $B^2_{s,s+1}(x)$ over $[-1, 1]$ are identical.

When we consider the more general case of minimizing $B^w_{s,t}$ with $w(x)$ as above, we have

$$X_1'X_2 = \begin{bmatrix} \mu_{s+1} & \mu_{s+2} & \cdots & \mu_t \\ & & & \\ \mu_{2s+1} & \mu_{2s+2} & \cdots & \mu_{s+t} \end{bmatrix},$$

and so from the matrix equation in Theorem 3.6 and the ones for minimax bias designs given above, it follows that the optimal designs for $B^w_{s,t}$ above satisfy the minimax bias

conditions.

Note also that the Chebyshev allocation is also optimal for both of the criteria under consideration. The above can be summarized as follows.

Theorem 3.7. The maximum bias and the $B^W$ criteria, $w(x) = \frac{1}{\pi}(1-x^2)^{-1/2}$, are equivalent, i.e., a design which is optimal with respect to one of those criteria is also optimal with respect to the other.

The Chebyshev allocation satisfy the particular conditions for minimization of both criteria.

Figures 1,2,3,4, and 5 show the graphs of the variance and bias functions of the min-average bias designs and the optimal designs with respect to the $B^W$ criterion considered above, i.e., minimax bias designs. Note the improvement achieved by the minimax bias designs at the extreme portions of the interval.

We mentioned in the preceding chapter properties of the designs which place an equal number of observations at the Chebyshev zeros, or the Chebyshev allocation. Since the zeros of other orthogonal polynomials do not have as many special properties, it seems appropriate to conclude this section with statement summarizing all the properties of the Chebyshev allocation.

The designs which place an equal number of observations

Figure 3.1. Graphs of $V[\hat{y}_L(x)]$. The variance functions of a linear polynomial model fitted using the designs indicated in the figure. The unit for the variance scale is $\sigma^2/N$

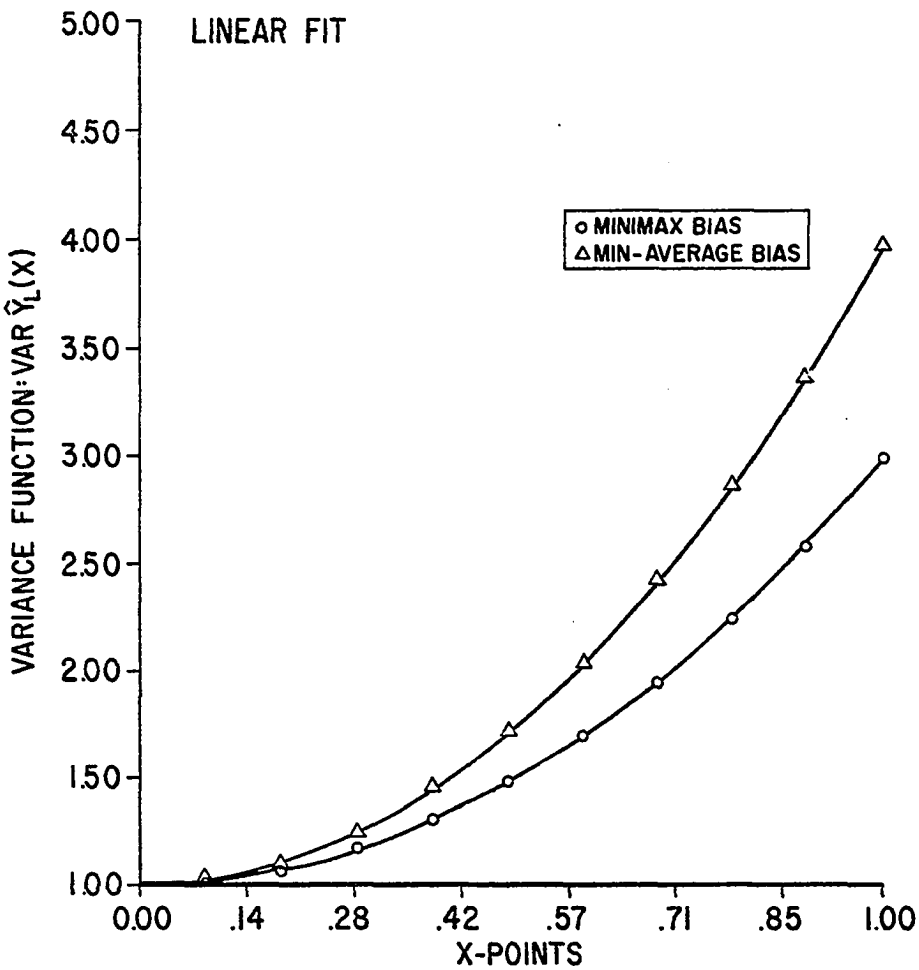

LINEAR FIT
MINIMAX BIAS
MIN-AVERAGE BIAS
VARIANCE FUNCTION: VAR $\hat{Y}_L(X)$
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.00
.14
.28
.42
.57
.71
.85
1.00
X-POINTS

Figure 3.2. Graphs of $V[\hat{y}_Q(x)]$. The variance functions of a quadratic polynomial model fitted using the designs indicated in the figure. The unit for the variance scale is $\sigma^2/N$

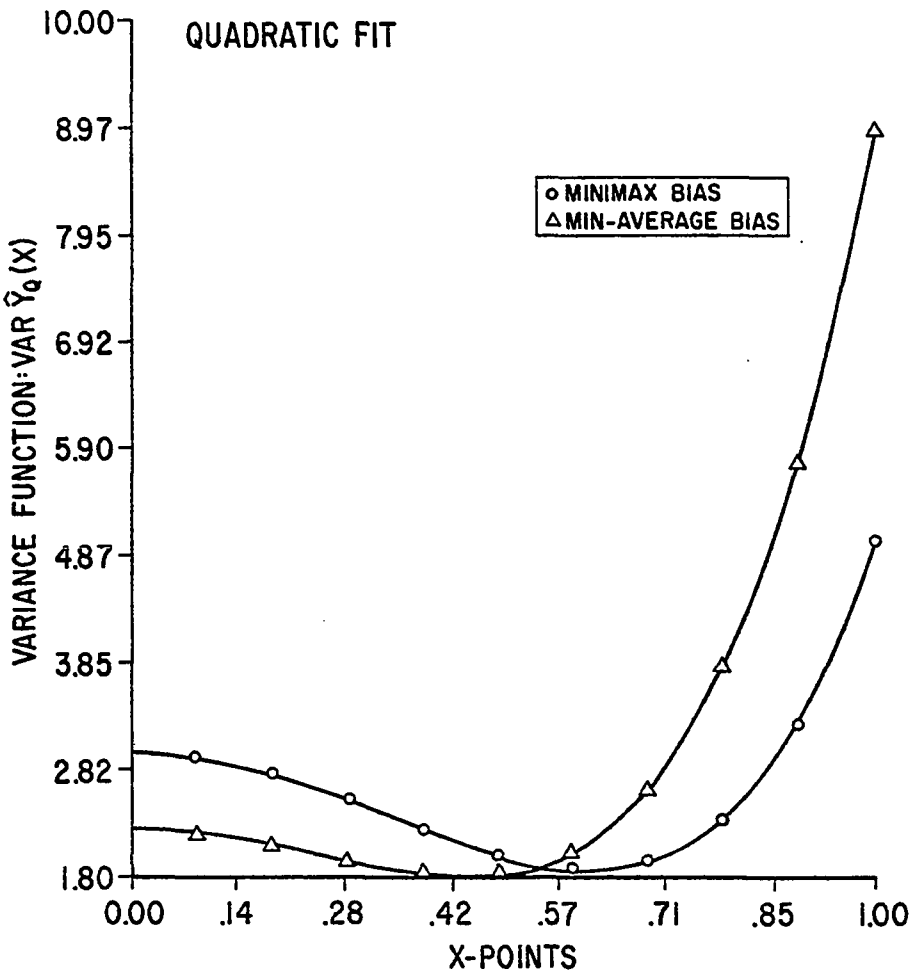

QUADRATIC FIT
o MINIMAX BIAS
Δ MIN-AVERAGE BIAS
VARIANCE FUNCTION: VAR $\hat{Y}_Q(X)$
10.00
8.97
7.95
6.92
5.90
4.87
3.85
2.82
1.80
0.00
.14
.28
.42
.57
.71
.85
1.00
X-POINTS

Figure 3.3. Graphs of $V[\hat{y}_C(x)]$. The variance functions of a cubic polynomial model fitted using the designs indicated in the figure. The unit for the variance sacle is $\sigma^2/N$

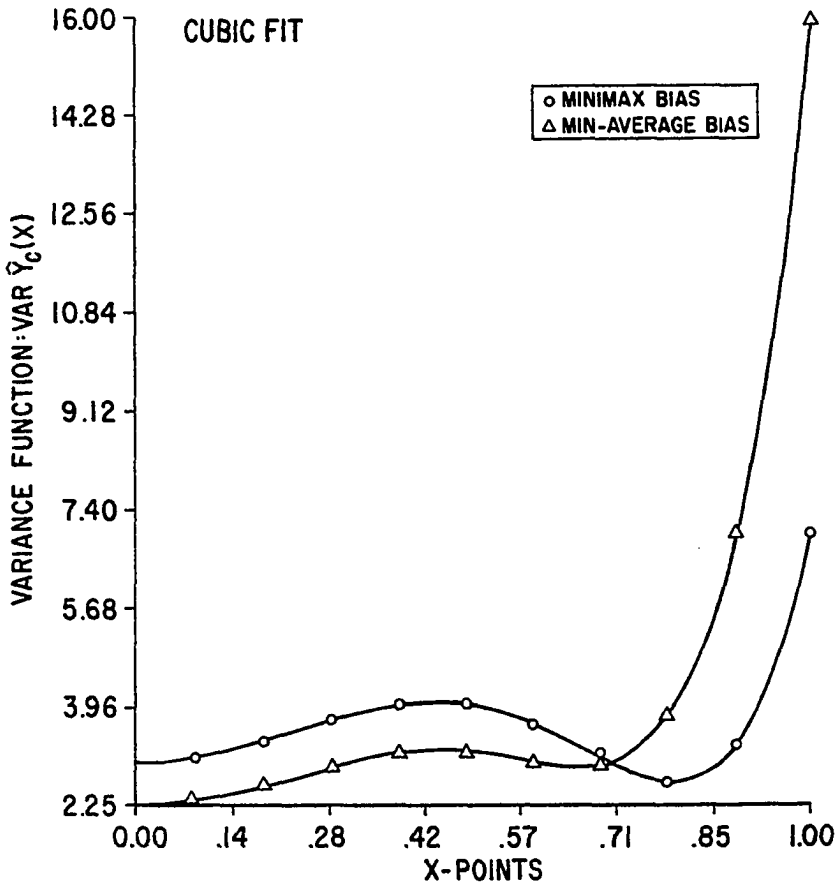

CUBIC FIT
o MINIMAX BIAS
Δ MIN-AVERAGE BIAS
VARIANCE FUNCTION: VAR $\hat{Y}_c(X)$
16.00
14.28
12.56
10.84
9.12
7.40
5.68
3.96
2.25
0.00
.14
.28
.42
.57
.71
.85
1.00
X-POINTS

Figure 3.4. Graphs of $B^2_{1,2}(x)$. The $(\text{bias})^2$ functions resulting from the fitting of a linear polynomial model using the designs indicated in the figure, when the true underlying model is a quadratic polynomial. The unit for the $(\text{bias})^2$ scale is $\beta_2^2$

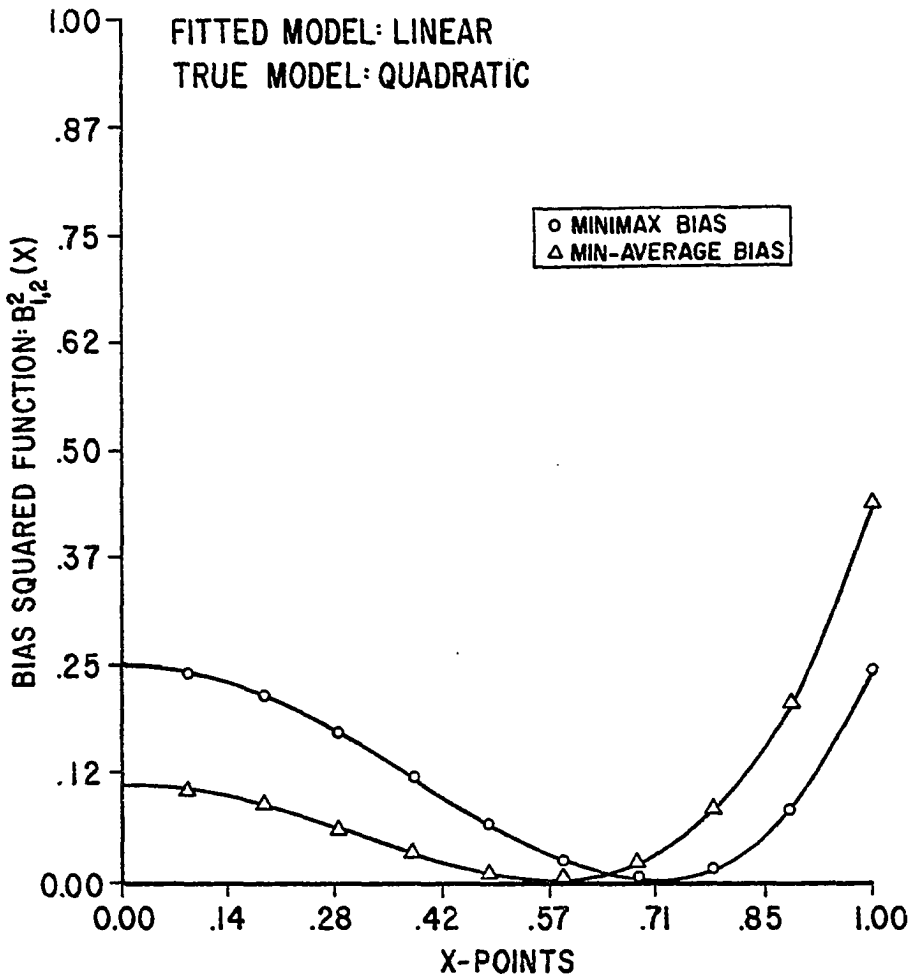
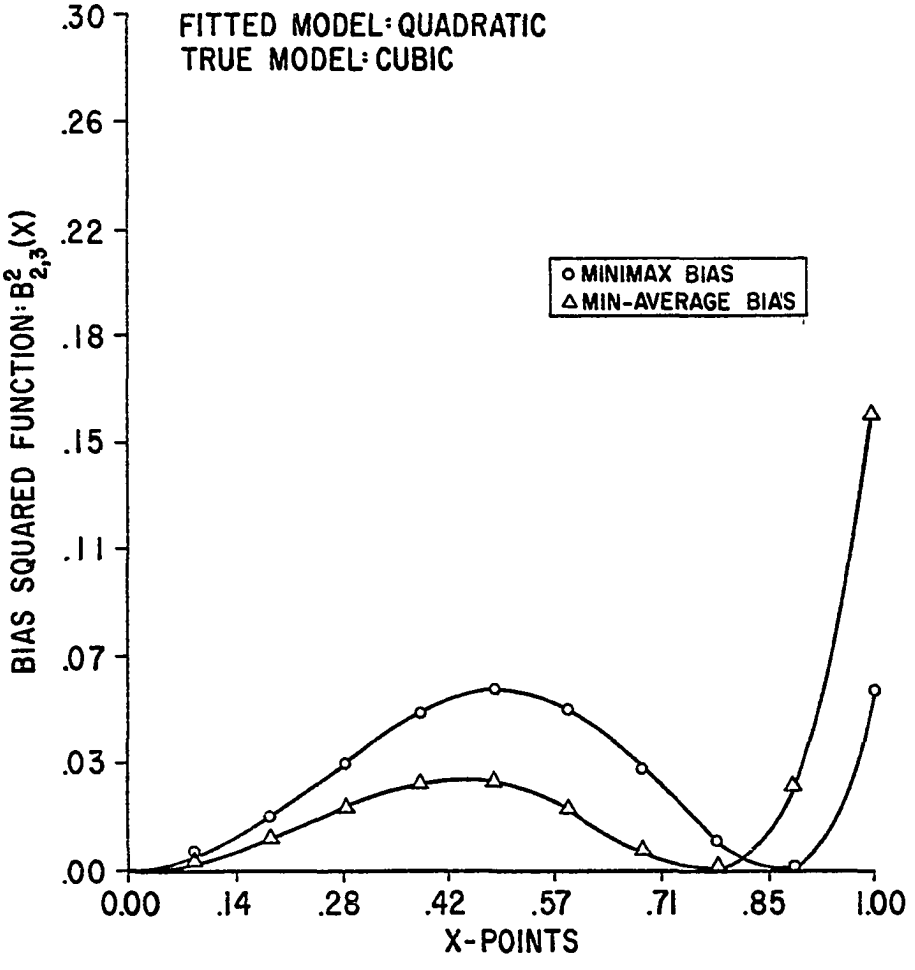

FITTED MODEL: LINEAR
TRUE MODEL: QUADRATIC
o MINIMAX BIAS
Δ MIN-AVERAGE BIAS
BIAS SQUARED FUNCTION: $B^2_{1,2}(X)$
1.00
.87
.75
.62
.50
.37
.25
.12
0.00
0.00
.14
.28
.42
.57
.71
.85
1.00
X-POINTS

Figure 3.5. Graphs of $B^2_{2,3}(x)$. The $(\text{bias})^2$ functions resulting from the fitting of a quadratic polynomial model with the designs indicated in the figure, when the true underlying model is a cubic polynomial. The unit for the $(\text{bias})^2$ scale is $\beta_3^2$


FITTED MODEL: QUADRATIC
TRUE MODEL: CUBIC
MINIMAX BIAS
MIN-AVERAGE BIAS
BIAS SQUARED FUNCTION: $B^2_{2,3}(X)$
.30
.26
.22
.18
.15
.11
.07
.03
.00
0.00
.14
.28
.42
.57
.71
.85
1.00
X-POINTS

at the zeros of a Chebyshev polynomial $T_n(x)$, with $n \geq [(s+t-1)/2]$ have the following properties:

(i) they minimize the mean average weighted square bias $B^w_{s,t}$ for $w(x) = (1-x^2)^{-1/2}$;

(ii) they minimize the maximum of the $(\text{bias})^2$ function $B^2_{n,n+1}(x)$ with $n$ as defined above;

(iii) they maintain the properties (i) and (ii) above in sequential least squares fitting of polynomial of degree $d < n$ when the true model has degree $d+1$;

(iv) they provide orthogonal least squares fitting if the polynomial fitted is expressed in terms of Chebyshev polynomials;

(v) in the case where the true underlying function is a general continuos function $f(x)$ with two derivatives, the corresponding least squares fitting converges uniformly to $f(x)$ in the interval of interest.

## 4. Another criterion of optimality

Let us now consider as criterion of optimality the mean average absolute bias, or

$$\int_{-1}^{1} |E\hat{y}_s(x) - \eta_t(x)| \, dx$$

where $\hat{y}_s(x)$ is the s-th degree fitted polynomial and $\eta_t(x)$ is the t-th degree true underlying polynomial model with $t > s$.

We will restrict ourselves to the determination of optimal

designs with respect to the above criterion for the case where $t = s+1$. As noted previously, without loss of generality we may consider a least squares $s$-th degree fit where we take observations at only $s+1$ distinct allocation points. The expression of the bias function is then

$$B_{s,s+1}(x) = |\pi_{s+1}(x)\ \beta_{s+1}|.$$

Note that the zeros of the $(s+1)$-th degree Chebyshev polynomial of second kind, $U_{s+1}(x)$, minimize the integral

$$\int_{-1}^{1} |\pi_{s+1}(x)|dx \quad ,$$

and its value is $2^{1-n}$ (Timan, 1963), so it follows from the above that the following result holds.

Theorem 3.8. Suppose a polynomial of degree $s$ is fitted and the true underlying model has degree $s+1$. A sufficient condition to minimize the average absolute bias

$$\int_{-1}^{1} |E\hat{y}_s(x)-\eta_{s+1}(x)|dx \ .$$

is to allocate an arbitrary number of observations at the zeros of the $(s+1)$-th degree Chebyshev polynomial of second kind, $U_{s+1}(x)$, the corresponding mean average absolute bias is $2^{1-s}|\beta_{s+1}|$.

# IV. COMPARISON BETWEEN OPTIMAL DESIGNS FOR REGRESSION

## A. Introduction

### 1. Aims

In this chapter we will compare the minimax bias designs with some of the optimal designs available in the literature. We will consider the designs which minimize the average square bias, or min-average bias designs, developed by Box and Draper (1959), the designs which minimize the maximum of the variance function over the interval of interest, or minimax variance designs, developed by Guest (1958), and the standard equal spacing designs. The behavior of the variance and $(\text{bias})^2$ functions over the interval of interest will be studied and comparisons will be made among the values of several criteria of optimality for each of the designs above.

Note that the min-average bias designs and the min-average bias estimable designs obtained in the last chapter differ only with respect to estimation of the average square bias. It follows that the bias and variance properties stated in this chapter for the min-average bias designs also hold for the min-average bias estimable designs.

## B. Variance Properties

### 1. Min-average bias designs

Guest (1958) obtained the limiting expression when the number of allocation points increases indefinitely, for the variance function of a polynomial fitted using an equal spacing design. Since as we will show in the following, the equal spacing designs converge to the min-average bias designs as the number of allocation points increases, the variance function obtained by Guest is exactly the variance function of a polynomial fitted using min-average bias design and it is given by

$$V[\hat{y}_s(x)] = [(s+1)\, P_s^2(x) - \frac{x^2-1}{s+1} P_s'^2(x)](s+1)\, \frac{\sigma^2}{N}, \qquad (4.1)$$

where $P_s(x)$ is the Legendre polynomial of degree $s$ defined over $[-1, 1]$.

Guest also proved that the variance function for the minimax variance fit

$$V[\hat{y}_s(x)] = [1+ \frac{x^2-1}{s(s+1)} P_s'^2(x)]\, (s+1)\, \frac{\sigma^2}{N} \qquad (4.2)$$

attains its maximum equal to $(s+1)\, \frac{\sigma^2}{N}$ at the end of the interval $[-1, 1]$ and at the zeros of $P_s'(x)$, the derivative of the Legendre polynomial of degree $s$. We can write the right-hand side of 4.1 in terms of the expression given by 4.2. Since the maximum of the absolute value of $P_s^2(x)$ is equal to 1 and is attained at the points $-1$ and 1, the above

leads to the following result:

Theorem 4.1. The maximum of the variance function of a s-th degree polynomial fitted using a min-average bias design is attained at the points $\pm 1$ and is given by $(s+1)^2 \frac{\sigma^2}{N}$ .

Let us now consider 4.1, and determine the average variance of a s-th degree polynomial fitted using a min-average bias design.

We have the following recurrence relation for Legendre polynomials (Sansone, 1959)

$$(x^2-1)\, P_s'(x) = sxP_s(x) - sP_{s-1}(x) ,$$

hence

$$\int (x^2-1)P_s'^2(x)\,dx = s\int_{-1}^{1} xP_s(x)P_s'(x)\,dx - p\int_{-1}^{1} P_{s-1}(x)P_s'(x)\,dx.$$

The following relationship also holds (Sansone, 1959)

$$\int_{-1}^{1} P_s(x)\, P_t(x)\, dx = \begin{cases} 0 , & (s\neq t) \\ \frac{2}{2s+1} , & (s=t) \end{cases}$$

$$P_s(-1) = (-1)^s, \quad P_s(1) = 1$$

and

$$P_s'(x) = (2s-1)\, P_{s-1} + (2s-5)\, P_{s-3}(x) + \ldots .$$

Thus, we have

$$\int_{-1}^{1} x\, P_s(x)\, P_s'(x)\, dx = \frac{2s}{2s+1}$$

and

$$\int_{-1}^{1} P_{s-1}(x)\, P_s'(x)\, dx = 2 .$$

It then follows that for the right-hand side of 4.1 above, we will have

$$\frac{1}{2} \int_{-1}^{1} \frac{(x^2-1)}{s+1}\, P_s'^2(x)\, dx = \frac{s}{2s+1}$$

and

$$\frac{s+1}{2} \int_{-1}^{1} P_s^2(x)\, dx = \frac{s+1}{2s+1} .$$

Adding the last two integrals we immediately obtain the following result:

Theorem 4.2. The average variance of a s-th degree polynomial fitted using a min-average bias design is given by $V_s = (s+1)\,\sigma^2/N$.

It follows from 4.1, that the variance functions for the linear, quadratic and cubic polynomials fitted using min-average bias designs are given respectively by:

$$V[\hat{y}_L(x)] = (1 + 3x^2)\frac{\sigma^2}{N} ,$$

$$V[\hat{y}_Q(x)] = \frac{9 - 18x^2 + 45x^4}{4}\frac{\sigma^2}{N} ,$$

and

$$V[\hat{y}_C(x)] = \frac{9 + 45x^2 - 165x^4 + 175x^6}{4}\frac{\sigma^2}{N} .$$

For the same fits above, from Theorem 4.1 we have:

$$\max_{-1\leq x\leq 1} V[\hat{y}_L(x)] = \frac{4\sigma^2}{N} , \quad \min_{-1\leq x\leq 1} V[\hat{y}_Q(x)] = \frac{9\sigma^2}{N} ,$$

and

$$\max_{-1\leq x\leq 1} V[\hat{y}_C(x)] = \frac{16\sigma^2}{N} .$$

Result 2 implies that:

$$V_L = \frac{2\sigma^2}{N} , \quad V_Q = \frac{3\sigma^2}{N} , \quad \text{and} \quad V_C = \frac{4\sigma^2}{N} .$$

where $V_L$, $V_Q$ and $V_C$ stand for the average variance of the linear quadratic, and cubic fits, respectively.

## 2. Equal spacing designs

We can write $K + 1$ points equally spaced in the $[-1\ 1]$ interval as

$$x_i = \frac{2i-K}{K} , \quad i = 0,1,\ldots,K .$$

It is then easy to prove that the first six moments of an equal spacing design are given by

$$\mu_1 = 0 \ , \quad \mu_2 = \frac{K+2}{3K} \ ,$$

$$\mu_3 = 0 \ , \quad \mu_4 = \frac{1}{15K^3}\,(3K^3 + 12K^2 + 8K - 8) \ ,$$

$$\mu_5 = 0 \ , \quad \mu_6 = (3K^6+21K^5+42K^4-56K^2+32)/(21K^5(K+1)) \ .$$

It is easy to see that equal spacing designs converge to min-average bias designs as the number of allocations increase. The limits of the design moments above, as $K$ increases, are the corresponding moments of a beta distribution $Be(1,1)$.

The variance functions of the linear quadratic and cubic polynomials fitted using equal spacing designs with $K+1$ allocation points are obtained by substituting the design moments above in the general expressions of the variance functions for symmetrical designs. These variance functions are given respectively by

$$V[\hat{y}_L(x)] = [1 - \frac{1}{\mu_2} x^2] \frac{\sigma^2}{N}$$

$$V[\hat{y}_Q(x)] = \frac{1}{(\mu_2\mu_4-\mu_2^3)} [\mu_2\mu_4+(\mu_4-3\mu_2^2)x^2+\mu_2 x^4] \frac{\sigma^2}{N}$$

$$V[\hat{y}_C(x)] = [(\frac{\mu_4}{\mu_4-\mu_2^2}) + (\frac{\mu_6}{\mu_2\mu_6-\mu_4^2} - \frac{2\mu_2}{\mu_4-\mu_2^2})x^2 \qquad (4.3)$$

$$+ (\frac{1}{\mu_4-\mu_2^2} - \frac{2\mu_4}{\mu_2-\mu_6-\mu_4^2})x^4 + (\frac{\mu_2}{\mu_2\mu_6-\mu_4^2})x^6] \frac{\sigma^2}{N} \ .$$

Similarly, we will obtain the average variances for the equal spacing designs, substituting their moments in the corresponding average variance general expressions

$$V_L = \left(1 + \frac{1}{3\mu_2}\right) \frac{\sigma^2}{N} ,$$

$$V_Q = \frac{1}{\mu_2\mu_4 - \mu_2^3} \left[15(\mu_2\mu_4 - \mu_2^3) - (3\mu_2 + 5\mu_4)\right] \frac{\sigma^2}{N} ,$$

and

$$V_C = \frac{1}{\Delta}\left[(\mu_2\mu_4\mu_6 - \mu_4^3) + \frac{1}{3}(2\mu_2\mu_4^2 + \mu_4\mu_6 - 3\mu_2^2\mu_6)\right.$$

$$\left. + \frac{1}{5}(2\mu_2^2\mu_4 + \mu_2\mu_6 - 3\mu_4^2) + \frac{1}{7}(\mu_2\mu_4 - \mu_2^3)\right]\frac{\sigma^2}{N}$$

where

$$\Delta = (\mu_4 - \mu_2^2)(\mu_2\mu_6 - \mu_4^2)$$

## 3. Minimax variance designs

The expression of the variance of a polynomial of degree $s$ fitted using a minimax variance design was obtained by Guest (1958) and is given by

$$V[\hat{y}_s(x)] = \left[1 + \frac{x^2 - 1}{s(s+1)} P_s'^2(x)\right] (s+1) \frac{\sigma^2}{N} \qquad (4.4)$$

where $P_s(x)$ is the Legendre polynomial of degree $s$ defined in the $[-1, 1]$ interval.

Guest also proved that the maximum of the variance function of a s-th degree polynomial fitted using minimax-variance design is given by $(s+1)\frac{\sigma^2}{N}$.

Let us compare the 4.1 and 4.2 which give the variance functions of s-th degree polynomials using min-average bias designs and minimax variance designs. From the value of the average variance for the min-average bias fit, we can easily obtain the corresponding value for the minimax variance designs and state the following result:

Theorem 4.3. The average variance of s-th degree polynomial fitted using a minimax variance design is $[2s(s+1)/(2s+1)]\sigma^2/N$.

From the above, we can write the variance functions of the linear, quadratic, and cubic polynomial fits using minimax variance designs as follows:

$$V[\hat{y}_L(x)] = (1 + x^2)\frac{\sigma^2}{N},$$

$$V[\hat{y}_Q(x)] = \frac{6 - 9x^2 + 9x^4}{2}\frac{\sigma^2}{N},$$

and

$$V[\hat{y}_C(x)] = \frac{13 + 33x^2 - 105x^4 + 75x^6}{4}\frac{\sigma^2}{N}.$$

The maxima of the variance functions are

$$\max_{-1\le x\le 1} V[\hat{y}_L(x)] = \frac{2\sigma^2}{N}, \quad \max_{-1\le x\le 1} V[\hat{y}_Q(x)] = \frac{3\sigma^2}{N},$$

and

$$\max_{-1\le x\le 1} V[\hat{y}_C(x)] = \frac{4\sigma^2}{N}.$$

From Theorem 4.3, the average variance values for the fits under consideration are given by

$$V_L = \frac{4}{3}\frac{\sigma^2}{N}, \quad V_Q = \frac{12}{5}\frac{\sigma^2}{N}, \quad \text{and} \quad V_C = \frac{24}{7}\frac{\sigma^2}{N}.$$

4. Minimax bias designs

From 4.3, the expressions of the variance functions for the linear, quadratic, and cubic polynomial fits using minimax bias designs are given by

$$V[\hat{y}_L(x)] = (1 + 2x^2)\frac{\sigma^2}{N},$$

$$V[\hat{y}_Q(x)] = (3 - 6x^2 + 8x^4)\frac{\sigma^2}{N},$$

and

$$V[\hat{y}_C(x)] = (3 + 12x^2 - 40x^4 + 32x^6)\frac{\sigma^2}{N}.$$

From Theorem 2.6, Chapter II, it follows that the maxima of the variance functions are

$$\max_{-1\le x\le 1} V[\hat{y}_L(x)] = \frac{3\sigma^2}{N}, \quad \max_{-1\le x\le 1} V[\hat{y}_Q(x)] = \frac{5\sigma^2}{N},$$

and

$$\max_{-1\le x\le 1} V[\hat{y}_C(x)] = \frac{7\sigma^2}{N}.$$

From the Theorem 2.7, the average variances have the following values:

$$V_L = \frac{5}{3}\frac{\sigma^2}{N}, \quad V_Q = \frac{13}{5}\frac{\sigma^2}{N}, \quad \text{and} \quad V_C = \frac{25}{7}\frac{\sigma^2}{N}$$

## C. Variance Comparisons

In this section we will compare the behavior as well as the average and maximum of the variance functions corresponding to the several designs considered above over the [-1, 1] interval. We will define the following measure of efficiency which we will also use for comparisons of bias.

Let d and d* be two different designs used to fit polynomials of the same degree. Designate by M(d) and M(d*) the values of some criterion which we are interested in minimizing, for designs d and d*.

We then define as the efficiency of design d with respect to design d*

$$e_{d,d^*} = \frac{M(d^*)}{M(d)} \quad .$$

If $e_{d,d^*} < 1$ then d* is said to be more efficient than d for the criterion under consideration.

### 1. Linear fit

Let MMB, MMV, MAB, E(4) and E(8) stand for minimax bias designs, minimax variance designs, min-average bias designs, and equal spacing designs with 4 and 8 points of allocation, respectively.

When compared with the minimax variance designs the efficiencies of the several designs with respect to the maximum and average of the variance function over the [-1, 1] interval are the following:

Min-average bias designs:

$$\max \{V[\hat{y}_L(x)\}: \quad e_{MAB,MMV} = .50,$$

and

$$\text{average } \{V[\hat{y}_L(x)]\}: \quad e_{MAB,MMV} = .67 .$$

Minimax bias designs:

$$\max \{V[\hat{y}_L(x)\}: \quad e_{MMB,MMV} = .66,$$

and

$$\text{average } \{V[\hat{y}_L(x)]\}: \quad e_{MMB,MMV} = .80.$$

Equal spacing designs:

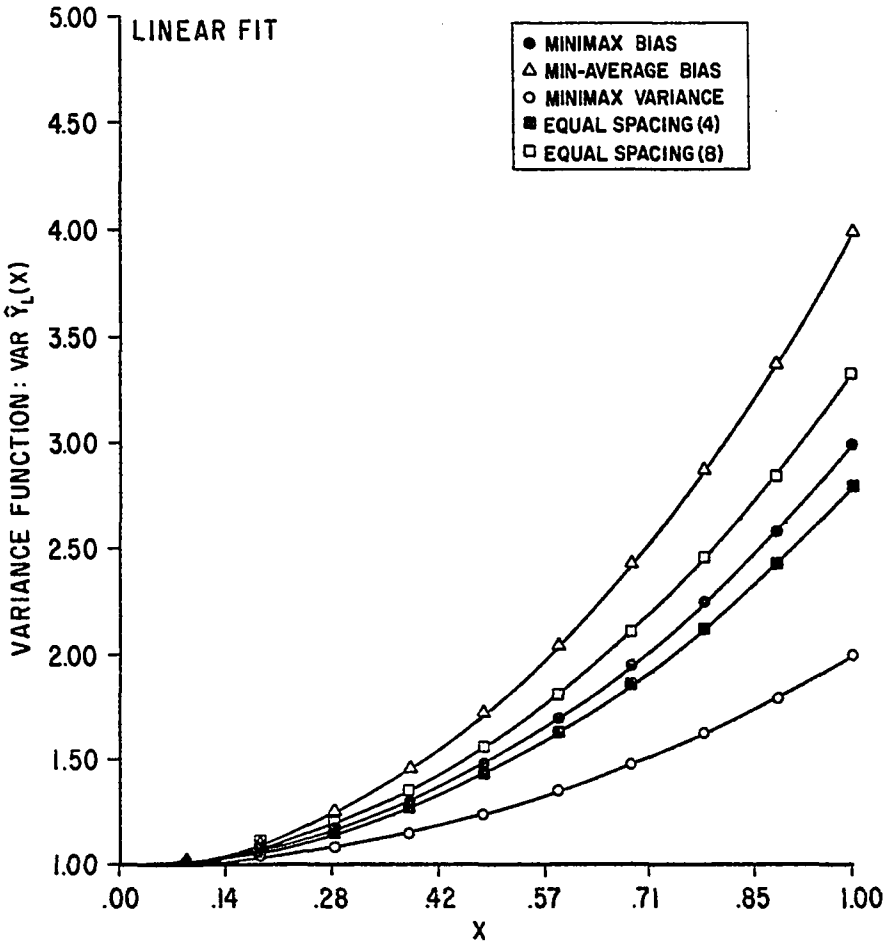
$$\max \{V[\hat{y}_L(x)]\}: \quad e_{E(4),MMV} = .71, \quad e_{E(8),MMV} = .60,$$

$$\text{average } \{V[\hat{y}_L(x)]\}: \quad e_{E(4),MMV} = .83, \quad \text{and}$$

$$e_{E(8),MMV} = .75 .$$

Figure 4.1 gives the graphs of the variance functions of the linear polynomial least squares fit for the designs we have been considering. We summarize in the following the conclusions concerning the behavior of the variance function and the efficiencies of these designs with regard to maxima and averages of their variance functions over the [-1, 1] interval.

Figure 4.1. Graphs of $V[\hat{y}_L(x)]$. The variance functions of a linear polynomial model fitted using the designs indicated in the figure. The unit for the variance scale is $\sigma^2/N$


LINEAR FIT
MINIMAX BIAS
MIN-AVERAGE BIAS
MINIMAX VARIANCE
EQUAL SPACING (4)
EQUAL SPACING (8)
VARIANCE FUNCTION: VAR $\hat{Y}_L(X)$
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.00
.14
.28
.42
.57
.71
.85
1.00
X

Minimax variance designs:

For a linear fit the minimax variance designs give a variance function that is uniformly smaller than the one of any other design. Its average variance function is also the smallest one.

Min-average bias designs:

Their variance function takes the largest values all over the interval. They also are the least efficient designs with respect to the maximum and average of the variance function.

Minimax bias designs:

Their variance function assumes reasonable values when compared with the variance functions of the other designs. The same can be said about efficiencies with respect to maximum and average over the $[-1, 1]$ interval.

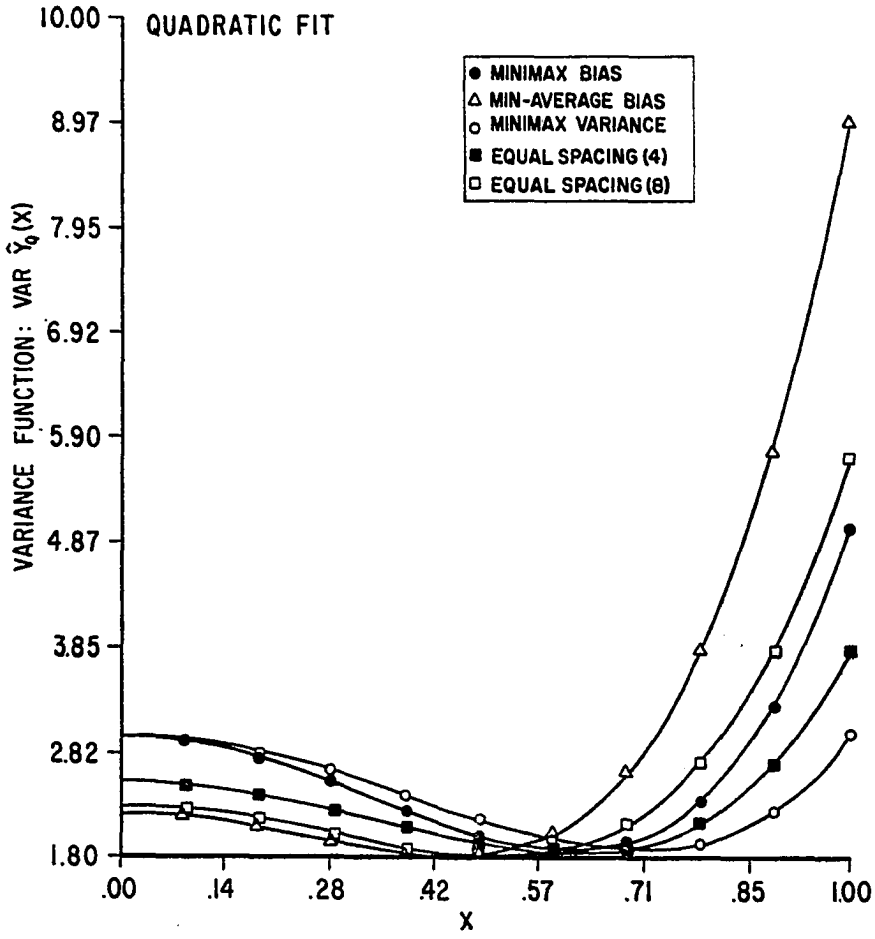
Equal spacing designs:

Their variance function approaches the one of the min-average bias designs as the number of allocations increase, so their behavior with regard to average and maximum value tends to that of the previous class.

2. Quadratic fit

Let us first consider efficiencies relative to the minimax variance design with respect to the maximum and average

Figure 4.2. Graphs of $V[\hat{y}_Q(x)]$. The variance functions of a quadratic polynomial model fitted using the designs indicated in the figure. The unit for the variance scale is $\sigma^2/N$


QUADRATIC FIT
MINIMAX BIAS
MIN-AVERAGE BIAS
MINIMAX VARIANCE
EQUAL SPACING (4)
EQUAL SPACING (8)
VARIANCE FUNCTION: VAR $\hat{Y}_Q(X)$
10.00
8.97
7.95
6.92
5.90
4.87
3.85
2.82
1.80
.00
.14
.28
.42
.57
.71
.85
1.00
X

of the variance function.

Minimax bias designs:

$$\max \{V[\hat{y}_Q(x)]\}: \quad e_{MMB,MMV} = .60,$$

and

$$\text{average } \{V[\hat{y}_Q(x)]\}: \quad e_{MMB,MMV} = .90$$

Min-average bias designs:

$$\max \{V[\hat{y}_Q(x)]\}: \quad e_{MAB,MMV} = .30,$$

and

$$\text{average } \{V[\hat{y}_Q(x)]\}: \quad e_{MAB,MMV} = .86$$

Equal spacing designs:

$$\max \{V[\hat{y}_Q(x)]\}: \quad e_{U(4),MMV} = .77, \quad e_{U(8),MMV} = .53,$$

$$\text{average } \{V[\hat{y}_Q(x)]\}: \quad e_{U(4),MMV} = 1.04,$$

and

$$e_{U(8),MMV} = .97 .$$

We now comment on the variance functions for the quadratic fits given by Figure 4.2 as well as on the list of efficiencies given above.

Minimax variance designs:

In the interval $|x| \leq .45$ the variance function of these designs takes reasonable values although they are the largest ones when compared with the values of the variance functions of the other designs. In extreme portions of the interval

(for $|x| > .70$) the minimax variance designs have the smallest variance function. Their average variance value is also smaller than that of any other of the designs under consideration.

Min-average bias designs:

For $|x| \leq .45$ the variance function of the min-average bias design takes smaller values than the variance function of any other design. For $|x| > .45$ the variance function of the min-average bias design is the one which takes the largest values, so its efficiency with regard to maximum of the variance function is very low. The average of the variance function is reasonable.

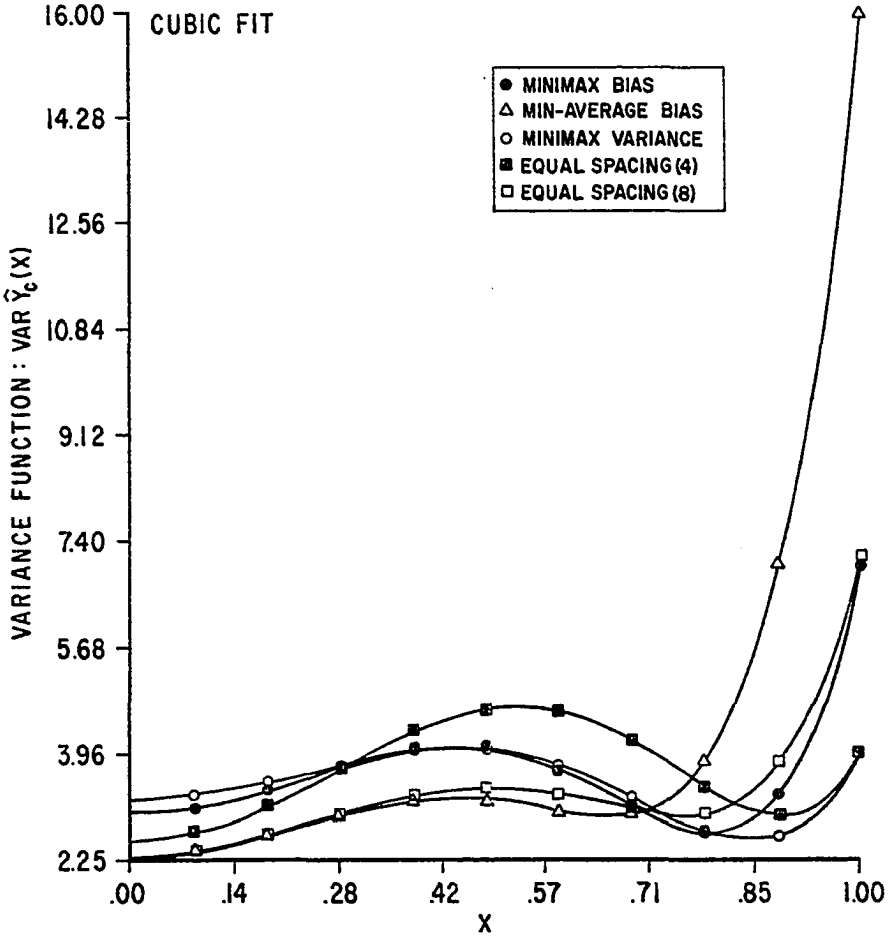
Minimax bias designs:

They have a variance function which assumes reasonable values all over the interval when compared with the variance function of any other design, and the same is true for the maximum and average of their variance function.

Equal spacing designs:

They have variance functions which are smaller than the variance function of the minimax variance designs over at least 50% of the interval, on its central part. The variance functions of the equal spacing designs take smaller values over the central part of the interval and larger values outside of it as the number of allocations increase and they

Figure 4.3. Graphs of $V[\hat{y}_C(x)]$. The variance functions of a cubic polynomial model fitted using the designs indicated in the figure. The unit for the variance scale is $\sigma^2/N$


CUBIC FIT
MINIMAX BIAS
MIN-AVERAGE BIAS
MINIMAX VARIANCE
EQUAL SPACING (4)
EQUAL SPACING (8)
VARIANCE FUNCTION : VAR $\hat{Y}_c(X)$
16.00
14.28
12.56
10.84
9.12
7.40
5.68
3.96
2.25
.00
.14
.28
.42
.57
.71
.85
1.00
X

converge to the min-average bias designs. The equal spacing design with 4 points of allocation is very efficient with regard to maximum and average of the variance function.

3. Cubic fit

We now examine the maximum and average variance efficiencies with respect to minimax variance designs for the case of cubic fit.

Minimax bias designs:

$\max \{V[\hat{y}_C(x)]\}$: $e_{MMB,MMV} = .57$,

and

average $\{V[\hat{y}_C(x)]\}$: $e_{MMB,MMV} = .96$

Min-average bias designs:

$\max \{V[\hat{y}_C(x)]\}$: $e_{MAB,MMV} = .25$,

and

average $\{V[\hat{y}_C(x)]\}$: $e_{MAB,MMV} = .86$

Equal spacing designs:

$\max \{V[\hat{y}_C(x)]\}$: $e_{E(4),MMV} = .84$, $e_{E(8),MMV} = .55$,

average $\{V[\hat{y}_C(x)]\}$: $e_{E(4),MMV} = .92$, and

$e_{E(8),MMV} = 1.05$.

Figure 4.3 shows the variance functions for the cubic fits of the designs we have been considering, and our comments about these and the corresponding maximum and average variance

efficiencies are as follows.

Minimax variance designs:

They have a variance function which is quite reasonable all over the interval, although in the central part for $|x| \leq .70$, it takes larger values than the variance function of the min-average bias designs. These designs also have smaller average variance than any of the other designs we are considering.

Min-average bias designs:

Their variance function is smaller than any other one for $|x| \leq .70$. Outside the $|x| \leq .80$ interval it assumes larger values than the variance function of any other designs. The efficiency with respect to the average value is high, but the variance function assumes very large values at the extreme portions of the interval.

Minimax bias designs:

The variance function of these designs is quite close to the one of the minimax variance designs for $|x| \leq .80$ and assumes larger values outside that interval. We can say that the variance function of the minimax bias designs has a good performance when compared with the variance function of any of the other designs.

Equal spacing designs:

For 8 allocation points the variance function of the equal spacing designs takes values near to those of the variance function of a min-average bias designs over the interval defined by $|x| \leq .70$. Outside that interval, the variance function of the equal spacing designs has better behavior than the variance of the min-average bias design; and it does not take as large values as the last one. With regard to maximum and average value the designs with 8 allocation points are very efficient.

General conclusions about variance properties of the linear, quadratic and cubic polynomial fits are as follows.

Minimax variance designs:

They have variance functions which assume quite reasonable values all over the interval of interest. They also have the smallest average variance values and, of course, minimize the maximum of the variance function over the interval of interest.

Min-average bias designs:

As noted earlier, for the linear fit these designs have variance function which assumes the largest values throughout the interval. For quadratic and cubic fits, the min-average bias designs have variance functions which take smaller values than the variance function of any other designs in the central part of the interval, but are the largest

ones at the extremes. To take advantage of the good variance properties of the min-average bias designs we need to know before-hand if the true model is quadratic or cubic and if the points where we want to predict are in the central part of the interval.

Minimax bias designs:

Their variance functions are fairly good all over the interval and as far as the maximum and average of the variance functions are concerned, they are more efficient than the min-average bias designs. The above implies that they are the more suitable for use in most of the fitting problems where we do not know before-hand the degree of the true underlying model nor the region where we would like to predict later.

Equal spacing designs:

If we restrict ourselves to approximately 8 allocation points in the quadratic and cubic cases, in the central part of the interval the equal spacing designs take as small variance values as the min-average bias designs. Outside, their variance functions take considerably smaller values than those of the min-average bias designs. It is interesting to note that for linear fit the equal spacing designs with 2 allocation points are also minimax variance designs, and equal spacing designs with 5 points are minimax bias designs.

In the quadratic case, the equal spacing designs with 3 allocation points are minimax variance designs. Also note that since the equal spacing designs converge to the min-average bias designs as the number of allocations increases, the behavior of their variance functions approaches the behavior of the variance functions of the min-average bias designs as the number of allocation points gets large.

## D. Bias Properties

### 1. Min-average bias designs

Let the fitted model be a linear polynomial and assume that the true underlying model is a quadratic model.

From Theorem 3.5 it follows that the $(\text{bias})^2$ function for the case above is given by

$$B^2_{1,2}(x) = [\tfrac{1}{3} - x^2]^2 \beta_2^2 .$$

If the model fitted is a quadratic model and the true polynomial is a cubic, also from Theorem 3.5 we have

$$B^2_{2,3}(x) = [\tfrac{3}{5} x - x^3]^2 \beta_3^2 .$$

From Theorem 3.5, the maxima of the $(\text{bias})^2$ function for the linear and quadratic fits considered above are attained at the points $-1$ and $+1$ and are given by

$$\max_{-1 \le x \le 1} B^2_{1,2}(x) = \tfrac{4}{9} \beta_2^2, \text{ and } \max_{-1 \le x \le 1} B^2_{2,3}(x) = \tfrac{4}{25} \beta_3^2 .$$

The mean average square bias for each of the cases above is also obtained from Theorem 3.5 and is given by

$$B_{1,2} = \frac{4}{45}\beta_2^2 \quad \text{and} \quad B_{2,3} = \frac{4}{175}\beta_3^2 .$$

Since the min-average bias designs are symmetrical, from Section A.3, Chapter III, it is seen that we will have

$$B_{1,3} = B_{1,2} + B_{2,3} .$$

Hence the average square bias when we use min-average bias designs to fit a linear polynomial and the true model is a cubic polynomial, is

$$B_{1,3} = \frac{4}{45}\beta_2^2 + \frac{4}{175}\beta_3^2 .$$

2. Equal spacing designs

We now study bias properties of the equal spacing designs.

The $(\text{bias})^2$ functions for the equal spacing designs will be obtained substituting the values of $\mu_2$ and $\mu_4$ in 2.2 and 2.4. The average square bias will be obtained proceeding similarly with 3.12 and 3.14.

3. Minimax variance designs

Let the fitted polynomial have degree $s$ and assume that the true underlying model has degree $s+1$. It there follows from Guest (1958) that the $(\text{bias})^2$ function is given by

$$B^2_{s,s+1}(x) = [(1-x^2)P'_s(x)]^2 \beta^2_{s+1} .$$

We have

$$\frac{d}{dx}[(1-x^2)P'_s(x)]^2 = (1-x^2)P''_s(x)-2xP'_s(x)$$

and since (Sansone, 1959)

$$(1-x^2)P''_s(x)-2x P'_s(x) = -n(n+1)P_s(x) ,$$

it follows that we have an interior stationary point at say $x_0$ if and only if $P_s(x_0) = 0$.

The relationship

$$(x^2-1)P'_s(x) = \frac{s(s+1)}{2s+1}[P_{s+1}(x)-P_{s-1}(x)] ,$$

holds for Legendre polynomials (Whittaker and Watson, 1934), and as stated in Section B.1 above, we also have

$$\int_{-1}^{1} P^2_s(x)dx = \frac{2}{2s+1} .$$

It then follows that

$$\frac{1}{2}\int_{-1}^{1}[(1-x^2) P'_s(x)]^2dx = \frac{2^{s+1}(s+1)^2[(s!)]^4}{(2s-1)(2s+1)(2s+3)(2s!)^2}$$

As an immediate consequence of the above we have the following result.

Theorem 4.4. Suppose a polynomial of degree $s$ is fitted using a minimax variance design, when the true model is of degree $s+1$. The $(\text{bias})^2$ function is

$$B^2_{s,s+1}(x) = [(1-x^2)\, P'_s(x)]^2\, \beta^2_{s+1} ,$$

and for $s > 1$ its maximum and minimum in the interior of $[-1, 1]$ occur among the zeros of $P_s(x)$.

The mean average bias is given by

$$B_{s,s+1} = \frac{2^{s+1}(s+1)^2[(s!)]^4}{(2s-1)(2s+1)(2s+3)(2s!)^2}\, \beta^2_{s+1} .$$

It follows from the above theorem that for the minimax variance designs we have

$$B^2_{1,2}(x) = [1-x^2]^2\, \beta^2_2 , \quad \text{and} \quad B_{2,3}(x) = [x-x^3]^2\, \beta^2_3 .$$

The maxima of $B^2_{1,2}(x)$ and $B^2_{2,3}(x)$ are attained respectively at the points $0$ and $\pm\sqrt{3/3}$ and given by

$$\max_{-1\le x\le 1} B^2_{1,2}(x) = \beta^2_2, \quad \text{and} \quad \max_{-1\le x\le 1} B^2_{2,3}(x) = \frac{4}{27}\, \beta^2_2 .$$

It also follows from Theorem 4.4 that

$$B_{1,2} = \frac{8}{15}\, \beta^2_2, \quad \text{and} \quad B_{2,3} = \frac{8}{105}\, \beta^2_3.$$

Since minimax variance designs are symmetrical, the mean average square bias from fitting a linear model when the true polynomial is a cubic is given by

$$B_{1,3} = \frac{8}{15}\beta_2^2 + \frac{8}{105}\beta_3^2 .$$

4. Minimax bias designs

We now consider the case of linear and quadratic polynomials fits when the true models are respectively quadratic and cubic.

It follows from Theorem 2.5 that

$$B_{1,2}^2(x) = [1/2-x^2]^2\,\beta_2^2, \quad B_{2,3}^2(x) = [\tfrac{3}{4}x-x^3]^2\,\beta_3^2,$$

$$\max_{-1\le x\le 1} B_{1,2}^2(x) = \frac{\beta_2^2}{4}, \quad \max_{-1\le x\le 1} B_{2,3}^2(x) = \frac{\beta_3^2}{16},$$

and that the mean average square bias are respectively

$$B_{1,2} = \frac{7}{60}\beta_2^2, \quad \text{and} \quad B_{2,3} = \frac{17}{560}\beta_3^2 .$$

Since the above designs are symmetric, we have

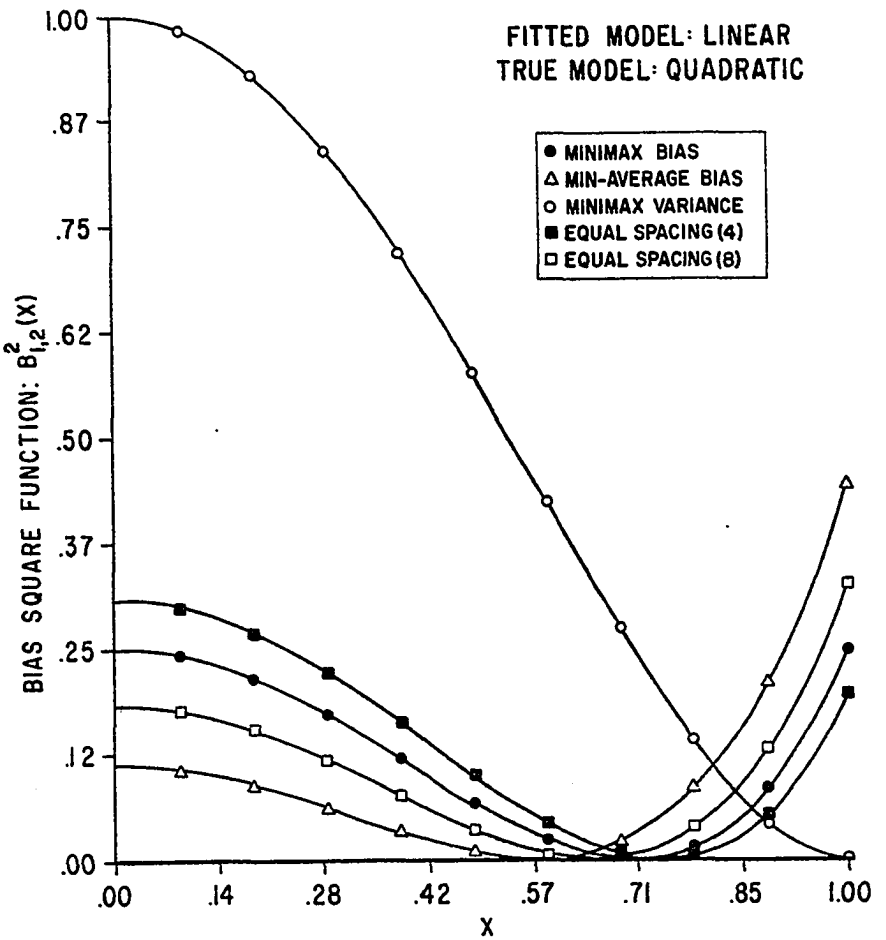
$$B_{1,3} = \frac{7}{60}\beta_2^2 + \frac{17}{560}\beta_3^2 .$$

## E. Bias Comparisons

1. Fitting a linear polynomial when the true model is a quadratic polynomial

We first compare the efficiency of the above designs with the minimax bias designs and min-average bias designs for the maximum and average value of the $(\text{bias})^2$ function. From the previous section we obtain the following values:

Figure 4.4. Graphs of $B^2_{1,2}(x)$. The $(\text{bias})^2$ functions resulting from the fitting of a linear polynomial model using the designs indicated in the figure, when the true underlying model is a quadratic polynomial. The unit for the $(\text{bias})^2$ scale is $\beta_2^2$


FITTED MODEL: LINEAR
TRUE MODEL: QUADRATIC
MINIMAX BIAS
MIN-AVERAGE BIAS
MINIMAX VARIANCE
EQUAL SPACING (4)
EQUAL SPACING (8)
BIAS SQUARE FUNCTION: $B^2_{1,2}(X)$
1.00
.87
.75
.62
.50
.37
.25
.12
.00
.00
.14
.28
.42
.57
.71
.85
1.00
X

Min-average bias designs:

$$\max_x \{B^2_{1,2}(x)\}: \quad e_{MAB,MMB} = .56.$$

Equal spacing designs:

$$\max \{B^2_{1,2}(x)\}: \quad e_{E(4),MMB} = .81, \quad e_{E(8),MMB} = .76$$

$$\text{average } \{B^2_{1,2}(x)\}: \quad e_{E(4),MAB} = .65, \text{ and } e_{E(8),MAB} = .91 .$$

Minimax variance designs:

$$\max \{B^2_{1,2}(x)\}: \quad e_{MMV,MMB} = .25$$

$$\text{average } \{B^2_{1,2}(x)\}: \quad e_{MMV,MAB} = .17 .$$

Minimax bias designs:

$$\text{average } \{B^2_{1,2}(x)\}: \quad e_{MMB,MAB} = .76 .$$

From Figure 4.4 which shows the graphs of the $(\text{bias})^2$ functions when the model fitted is linear and the true model is quadratic, as well as from the list of efficiencies given above, the following conclusions can be reached about the designs under consideration.

Minimax variance designs:

They have a $(\text{bias})^2$ function, which assumes very large values over 80% of the interval. They are the least efficient in this case.

Min-average bias designs:

Their $(\text{bias})^2$ function takes values smaller than any other one for $|x| \leq .58$, but it is very large outside of the above interval when compared with the $(\text{bias})^2$ of other designs. As in the case of the variance function, the $(\text{bias})^2$ function assumes very large values in the extreme portions of the interval.

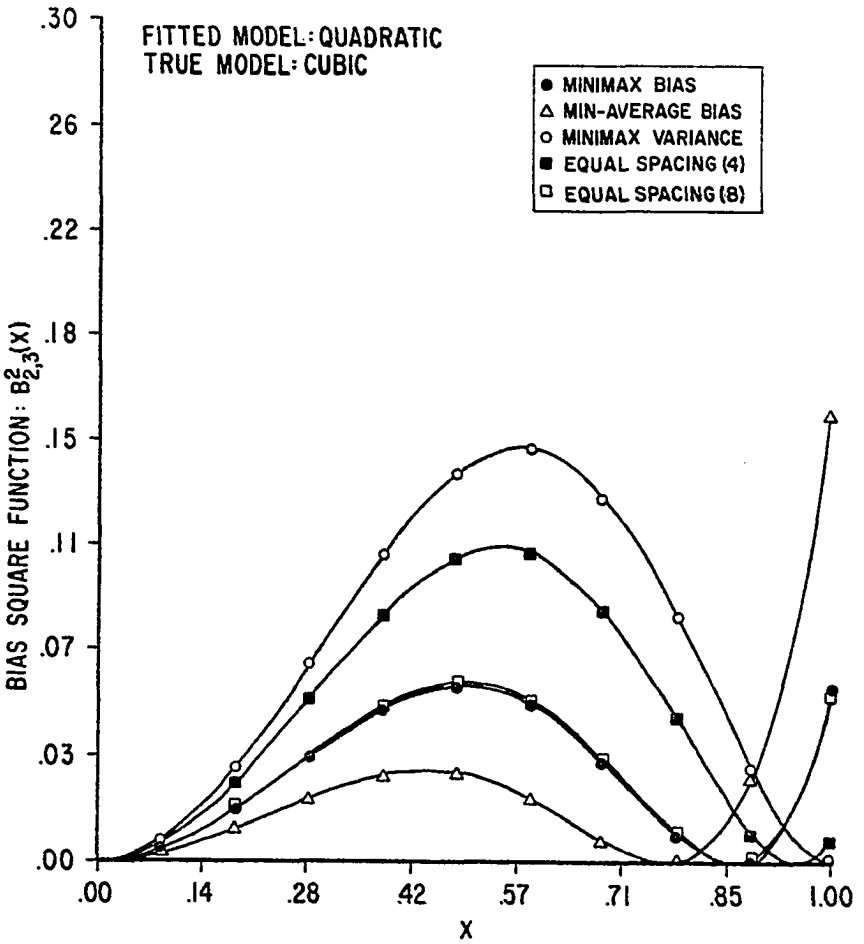
Minimax bias designs:

They have a $(\text{bias})^2$ function with a quite reasonable behavior over the whole interval. It does not assume as small values as the $(\text{bias})^2$ function of the min-average bias designs over the central part of the interval but it is better outside. Their average bias efficiency is reasonably high.

Equal spacing designs:

They have $(\text{bias})^2$ functions which assume larger values than the $(\text{bias})^2$ function of the min-average bias designs in the central part of the interval, but smaller values outside. For the reasons mentioned earlier, the behavior of the $(\text{bias})^2$ functions of the equal spacing designs approach that of the min-average bias designs as the number of allocation points increases.

Figure 4.5. Graphs of $B^2_{2,3}(x)$. The $(\text{bias})^2$ functions resulting from the fitting of a quadratic polynomial model with the designs indicated in the figure, when the true underlying model is a cubic polynomial. The unit for the $(\text{bias})^2$ scale is $\beta_3^2$


FITTED MODEL: QUADRATIC
TRUE MODEL: CUBIC
MINIMAX BIAS
MIN-AVERAGE BIAS
MINIMAX VARIANCE
EQUAL SPACING (4)
EQUAL SPACING (8)
BIAS SQUARE FUNCTION: $B^2_{2,3}(X)$
.30
26
.22
.18
.15
.11
.07
.03
.00
.00
.14
.28
.42
.57
.71
.85
1.00
X

2. Fitting a quadratic when the true model is a cubic

A list of efficiency comparisons similar to the one considered in the linear case follows:

Min-average bias designs:

$$\max \{B^2_{2,3}(x)\}: \quad e_{MAB,MMB} = .39 .$$

Equal spacing designs:

$$\max \{B^2_{2,3}(x)\}: \quad e_{E(4),MMB} = .45, \quad e_{E(8),MMB} = .98,$$

$$\text{average } \{B^2_{2,3}(x)\}: \quad e_{E(4),MAB} = .30, \text{ and}$$

$$e_{E(8),MAB} = .78 .$$

Minimax variance designs:

$$\max \{B^2_{1,2}(x)\}: \quad e_{MMV,MMB} = .42,$$

and

$$\text{average } \{B^2_{1,2}(x)\}: \quad e_{MMV,MAB} = .30 .$$

Minimax bias designs:

$$\text{average } \{B^2_{2,3}(x)\}: \quad e_{MMB,MAB} = .75 .$$

The list of efficiencies above and the graphs in Figure 4.5 lead us to the following remarks.

Minimax bias designs are the ones with better variance properties on the average, and the min-average bias designs are suitable for fitting when predictions are to be made at

the center of the interval. Equal spacing designs are close to minimax bias designs when the number of allocation points is small. In this case the best equal spacing design has 8 distinct allocation points. As the number of allocation points increases the equal spacing designs converge to the min-average bias designs. The bias properties of these designs present the same patterns as in the previous case.

3. Final conclusions

We summarize in Table 4.1 the general values obtained for the several criteria involved in this thesis for each of the designs considered.

From the behavior of the variance and bias functions as well as from Tables 4.1 and 4.2, and by combining the variance and bias properties of the several designs, our conclusions are as follows:

Minimax variance designs:

They have good variance related properties and so are efficient when the true underlying model is known before experimentation. Otherwise, the minimax variance designs are very inefficient. They are very sensitive to departures from the true model.

Table 4.1. Expressions for values of the various criteria in the designs considered

| Design \ Measure of goodness | Average Var. | Average Bias | Max. Bias | Max. Var. |
|---|---|---|---|---|
| Minimax var. | $\frac{2s(s+1)}{2s+1}\frac{\sigma^2}{N}$ * | $\frac{2^{s+1}(s+1)^2[(s!)]^4}{(2s-1)(2s+1)(2s+3)(2s!)^2}\beta_{s+1}^2$ | | $(s+1)\frac{\sigma^2}{N}$ |
| Min-avg. bias | $(s+1)\frac{\sigma^2}{N}$ | $\frac{1}{2s+3}[\frac{2^{s+1}(s+1)!}{(2s+2)!}]^2\beta_{s+1}^2$ | $[\frac{2^{s+1}[(s+1)!]}{(2s+2)!}]^2\beta_{s+1}^2$ | $(s+1)^2\frac{\sigma^2}{N}$ |
| Minimax bias | $\frac{2s^2+2s+1}{2s+1}\frac{\sigma^2}{N}$ | $4^{-(s+1)}\frac{2s^2+4s+1}{4s^2+8s+3}$ | $4^{-s}\beta_{s+1}^2$ | $(2s+1)^2\frac{\sigma^2}{N}$ |

*Obtained by Guest (1958).

Table 4.2. Designs efficiencies for linear, quadratic and cubic fits

| Efficiencies / Measure of Efficiency | Linear Fit | | | | Quadratic Fit | | | | Cubic Fit | |
|---|---|---|---|---|---|---|---|---|---|---|
| Designs | Max. Bias | Avg. Bias | Max. Var. | Avg. Var. | Max. Bias | Avg. Bias | Max. Var. | Avg. Var. | Max. Var. | Avg. Var. |
| Minimax bias | 1.00 | .76 | .66 | .80 | 1.00 | .75 | .60 | .90 | .57 | .96 |
| Min-avg. bias | .56 | 1.00 | .50 | .67 | .39 | 1.00 | .42 | .86 | .26 | .85 |
| Minimax var. | .25 | .17 | 1.00 | 1.00 | .42 | .30 | 1.00 | 1.00 | 1.00 | 1.00 |
| Equal Sp. (4) | .81 | .76 | .71 | .83 | .45 | .30 | .77 | 1.04 | .84 | .92 |
| Equal Sp. (8) | .65 | .91 | .60 | .75 | .98 | .78 | .53 | .97 | .55 | 1.05 |

Min-average bias designs:

The variance function for the linear polynomial fit is very large. For the quadratic and cubic fits, the variance and $(\text{bias})^2$ functions have very similar behavior. They are the best ones in the central part of the interval and the worst outside. Thus, to use these designs efficiently we should know beforehand if the true model is quadratic or cubic and if the region for which we wish to make predictions lies in the central part of the interval or not.

Equal spacing designs:

When the number of allocation points is between 6 and 10, those designs have $(\text{bias})^2$ and variance functions quite close to the ones for the minimax bias designs. Since they converge to the min-average bias designs, as the number of allocations points increases, they display similar properties when the number of allocations is bigger than 10.

Minimax bias designs:

As we have seen above, these designs are good as far as both variance and bias properties. The use of the Chebyshev allocation allows orthogonal fitting and in accordance with Section C, of Chapter II they should give good approximations even if the true underlying model is not a polynomial. These comments lead us to conclude that the minimax bias designs are the most suitable for most practical applications.

We finally notice that while min-average bias designs and minimax bias designs have fairly good variance properties, minimax variance designs are quite inefficient with respect to bias. The above leads us to infer that usually optimal designs for bias related criteria are also good for situations where there is both bias and variance. A similar conclusion was reached by Box and Draper (1959) in their consideration of first degree polynomial models.

# V. RANDOM STAR DESIGNS

## A. Introduction

### 1. Factorial experiments

Much experimentation is concerned with the effects of categorical factors on a yield or response variable. A simple illustrative example is the effect on performance of students at the end of a course on a defined body of knowledge. This performance can be measured by some test which has been carefully constructed. It is clear that performance is influenced by the text-book used, the instructor, the instruction method, the background of student, and so on. There is indeed, a very large number of potentially relevant factors and many of these are categorical or qualitative.

The whole theory of factorial designs has been directed at the formulation of useful plans of observation for such situations. There is a wide spectrum of situations ranging from those in which there is a small number of factors which one attempts to study in full detail to those in which the number of factors thought to be possibly relevant is large and one makes an exploratory study. In all cases, the designer of the experiment must make a choice between examining a few of the possible factorial combinations very carefully and deeply or examining a large number of combinations with small effort on each combination. A further extension of this notion is the use of fractional replication in which only a selected

subset of the possible combinations is examined. Basic ideas of this procedure are given by Kempthorne (1952).

The notion of fractional replication in broad terms is that instead of examining all M, say, of the combinations of factors one looks at a subset of size $M_i$, say, of these. The role of fractional replication is discussed by Kempthorne (1952). The subset of combinations used in the experiment is chosen from the totality on the basis of assumptions of the magnitudes of effects and interactions. If, for instance, it is thought that there are negligible interactions, a main effect plan will be used (Kempthorne, (1952) and Addelman and Kempthorne (1961)).

To describe the situation, it is necessary to give some notation. We suppose that we have n factors, say $F^{(1)}$, $F^{(2)},\ldots,F^{(n)}$ with levels denoted by $x_1,x_2,\ldots,x_n$.

If the factor $F^{(i)}$ has $n_i$ levels, the possible values for $x_i$ are usually taken to be $1,2,\ldots,n_i$ or $0,1,2,\ldots,n_i-1$. While these are integers they are really only labels of the possibilities and do not have an arithmetic nature. Then we can envisage that there is a true yield, say $\eta(x_1,x_2,\ldots,x_n)$ associated with the combination level $x_1$ of factor $F^{(1)}$, level $x_2$ of factor $F^{(2)}$, and so on, to level $x_n$ of factor $F^{(n)}$. We can also assume, unless we have reason to suppose otherwise, that an observation on the combination $x = (x_1,x_2,\ldots,x_n)$, say, $y(x_1,x_2,\ldots,x_n)$ is representable by

the observational equation

$$y(x_1,x_2,\ldots,x_n) = \eta(x_1,x_2,\ldots,x_n),$$

or when we assume the existence of an additive error term, by

$$y(x_1,x_2,\ldots,x_n) = \eta(x_1,x_2,\ldots,x_n) + e(x_1,x_2,\ldots,x_n)$$

where $e(x_1,x_2,\ldots,x_n)$ is an error of observation. Usually, we assume that the e's for different vectors $(x_1,x_2,\ldots,x_n)$ or different observations on the same vector are uncorrelated with mean zero and common variance $\sigma^2$.

A main effect plan is one which is aimed at the examination of the main effects under the assumption that there are no interactions. Formally, this is representable by the relation

$$\eta(x_1,x_2,\ldots,x_n) = \mu + \sum_{i=1}^{n} E^{(i)}_{x_i}$$

in which $E^{(i)}_{x_i}$ for $x_i = 1,2,\ldots,n_i$ are additive contributions associated with the different levels of the factor $F^{(i)}$. A two-factor-interaction design is one which allows the estimation, subject to the standard non-estimability features of a classificatory linear model of the relationship, of the model

$$\eta(x_1,x_2,\ldots,x_n) = \mu + \sum_{i=1}^{n} E^{(i)}_{x_i} + \sum_{\substack{i,j=1 \\ i<j}}^{n} (E^{(i)}E^{(j)})_{x_i x_j} .$$

The potential aims of factorial experiments include factor screening, that is the development of ideas of whether factors have a negligible role in explaining $\eta(x_1,x_2,\ldots,x_n)$. They also include estimations of responses associated with some or all of the possible vectors $(x_1,x_2,\ldots,x_n)$.

The aim of studying the nature of the function $\eta(x_1,x_2,\ldots,x_n)$ for the points $(x_1,x_2,\ldots,x_n)$ of the full factorial set is the one with which we shall be concerned. This function exists just for the full set of factorial combinations, which will be a set of $\prod_{i=1}^{n} L_i$ points on a lattice in n dimensions.

The case is interesting because of the analogy to the problem discussed in the first part of this thesis. In that part, we saw that it was useful and perhaps necessary to introduce a weighting function $w(x_1,x_2,\ldots,x_n)$ which reflects the interest of the experimenter in the parts of the whole experimental region. In the case of most qualitative factorial experiments, it seems quite compelling to take the view that interest of the experimenter will often be in all the possible $\prod_{i=1}^{n} n_i$ points in the factorial space, because there is usually no natural ordering of the levels of categorical factors. So one aim of the experimenter will often be to form an idea, e.g., to obtain an estimate of the function $\eta(x_1,x_2,\ldots,x_n)$ over the lattice of points, with equal weight associated with every point.

We now proceed by presenting the relationship between the concepts and ideas described above and the general aims of this chapter.

## 2. Aims

The problem of forming some conclusions about a population of $N$ individuals where one is able to investigate $m$, less than $N$, is a standard one of statistical practice. A solution which is widely, but not completely, accepted is to select the $m$ individuals at random from the population of $N$ individuals. In the context of factorial experiments, this amounts to a random selection of a subset of treatment combinations from the full factorial system. One then estimates the unobserved points of the lattice by the fitting of some statistical model. We will consider the case where the points observed and used for estimation are the elements of what we call a "star design" which we will define. We will be interested in obtaining the expression of the average mean square error of prediction of the numbers $\eta(x_1,x_2,\ldots,x_n)$ for all points of the lattice, and the actual set observed is a random one from a class of sets. The final aim will be the overall operating characteristics of the random sampling procedure.

## B. Random Star Designs

### 1. Statement of the problem

Consider the case where we have $n$ factors $F^{(1)}, F^{(2)}, \ldots, F^{(n)}$ with respectively $L_1, L_2, \ldots, L_n$ levels and we denote the true yield at the point $(x_1, x_2, \ldots, x_n)$ by $\eta(x_1, x_2, \ldots, x_n)$. We define a star design as follows:

(i) Choose at random one level of each factor, and denote the selected levels by $a_1, a_2, \ldots, a_n$;

(ii) Observe the set of points

$$D\{(a_1, a_2, \ldots, a_n), (x_1, a_2, \ldots, x_n), (a_1, x_2, \ldots, a_n), \ldots,$$

$$(a_1, a_2, \ldots, x_n) \text{ with } x_i = 1, 2, \ldots, L_i, \ x_i \neq a_i\}. \quad (5.1)$$

The point $(a_1, a_2, \ldots, a_n)$ is said to be the center of the set and the observed points consist of all points on lines parallel to the edges of the lattice which pass through the center. The random feature of the design is that the center is chosen at random from the totality of points of the lattice. So, if we use $L$ to denote $L_1 L_2 \ldots L_n$ there are $L$ possible realizations of the design. Each realization consists of $n$ points where $m = \sum_{i=1}^{n} L_i - m + 1$. This design is of interest because it can be used for any completely crossed factorial structure, in contrast to fractional replication designs which are possible only with some symmetries. Furthermore, the star

designs permit the fitting of a main effect model. So the overall procedure is to pick a star set, fit the main effect model and then use the fitted model to estimate every point of the lattice. Our final interest is the average mean square error of estimation of the numbers $\eta(x_1,x_2,\ldots,x_n)$ for all points of the lattice. The procedure used will be to consider error conditional on the random center, and then average this over all possible centers.

## 2. Illustrative example

To illustrate the ideas contained in this chapter, consider the special case where we have three factors $F^{(1)}, F^{(2)}$ and $F^{(3)}$ with respectively $L_1, L_2$ and $L_3$ levels. A true yield at the lattice point $(x_1,x_2,x_3)$ can be written identically as

$$\eta(x_1,x_2,x_3) = \mu + E^{(1)}_{x_1} + E^{(2)}_{x_2} + E^{(3)}_{x_3} + (E^{(1)}E^{(2)})_{x_1x_2} + (E^{(1)}E^{(3)})_{x_1x_3} + (E^{(2)}E^{(3)})_{x_2x_3} + (E^{(1)}E^{(2)}E^{(3)})_{x_1x_2x_3} . \qquad (5.2)$$

where $E^{(i)}_{x_i}$, $(E^{(i)}E^{(j)})_{x_ix_j}$ etc. are the standard expressions for effects and interactions. Effects and interactions in general can be specified in several ways, all somewhat tedious (Zyskind, 1962 and White, 1963). For present purposes we give a highly abbreviated representation. Consider the average of the $\eta(x_1,x_2,\ldots,x_n)$ for which

$x_{r_1} = u_1$, $x_{r_2} = u_2, \ldots, x_{r_k} = u_k$ averaging being over all other x-values, and denote it $\eta(*,*,\ldots,u_1,*,u_2,*,*,u_r,\ldots)$ in which $u_1$ occurs in the $r_1$-th position, $u_2$ in the $r_2$-th position etc., and all other positions are occupied by *'s. This is an admissible mean. Now consider all admissible means which differ from this one by averaging over one or more of the non-averaged coordinates. Then finally form the linear combination of these with coefficient $(-1)^s$, where $s$ is the number of additional coordinates averaged over. The resultant number is denoted by $(E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{u_1u_2\ldots u_k}$.

We now return to the example. Let the model assumed for expressing the yields as function of the factor levels be

$$y(x_1,x_2,x_3) = \mu + E^{(1)}_{x_1} + E^{(2)}_{x_2} + E^{(3)}_{x_3},$$

with

$$x_1 = 1,2,\ldots,L_1, \quad x_2 = 1,2,\ldots,L_2 \text{ and } x_3 = 1,2,\ldots,L_3 .$$

It then follows from the above that

$$P\{x_1=a_1\} = \frac{1}{L_1} , \quad P\{x_2=a_2\} = \frac{1}{L_2} , \quad P\{x_3=a_3\} = \frac{1}{L_3} ,$$

and so for any triple $(x_1,x_2,x_3)$

$$P\{(x_1,x_2,x_3) = (a_1,a_2,a_3)\} = \frac{1}{L} .$$

Also note that the star with center $(a_1,a_2,a_3)$ is made up of the set of points defined by:

$$D(a_1,a_2,a_3) = \{(x_1,x_2,x_3):(a_1,a_2,a_3),\ (x_1,a_2,a_3),\ (a_1,x_2,a_3),$$

$$(a_1,a_2,x_3),\ \text{with}\ x_i = 1,2,\ldots,L_i,\ x_i \neq a_i\}.$$

Since the star designs are "saturated designs" we will now give the corresponding definition.

A design-model is saturated if its rank is equal to the maximal rank for the model, and there is no linear function of the observations with zero expectation.

We consider first the case for which there is no error of observation. It follows that for any observed point we have

$$y(x_1,x_2,\ldots,x_n) = \eta(x_1,\ldots,x_n).$$

Under the assumed model, we have

$$(E^{(1)}_{x_1} - E^{(1)}_{a_1}) = (\mu + E^{(1)}_{x_1} + E^{(2)}_{a_2} + E^{(3)}_{a_3}) - (\mu + E^{(1)}_{a_1} + E^{(2)}_{a_2} + E^{(3)}_{a_3})$$

$$(E^{(2)}_{x_2} - E^{(2)}_{a_2}) = (\mu + E^{(1)}_{a_1} + E^{(2)}_{x_2} + E^{(e)}_{a_3}) - (\mu + E^{(1)}_{a_1} + E^{(2)}_{a_2} + E^{(3)}_{a_3})$$

and

$$(E^{(3)}_{x_3} - E^{(3)}_{a_3}) = (\mu + E^{(1)}_{a_1} + E^{(2)}_{a_2} + E^{(3)}_{x_3}) - (\mu + E^{(1)}_{a_1} + E^{(2)}_{a_2} + E^{(3)}_{a_3}),$$

furthermore, the main effect model is saturated so the estimates for the parametric functions above are given by

$$(\hat{E}^{(1)}_{x_1} - E^{(1)}_{a_1}) = y(x_1,a_2,a_3) - y(a_1,a_2,a_3),$$

$$(\hat{E}^{(2)}_{x_2} - E^{(2)}_{a_2}) = y(a_1,x_2,a_3) - y(a_1,a_2,a_3),$$

and

$$(\hat{E}^{(3)}_{x_3} - E^{(3)}_{a_3}) = y(a_1,a_2,x_3) - y(a_1,a_2,a_3) .$$

Under the fitted model,

$$\eta(x_1,x_2,x_3) = (\mu + E^{(1)}_{a_1} + E^{(2)}_{a_2} + E^{(3)}_{a_3}) + (E^{(1)}_{x_1} - E^{(1)}_{a_1} + (E^{(2)}_{x_2} - E^{(2)}_{a_2}) + (E^{(3)}_{x_3} - E^{(3)}_{a_3}),$$

hence the estimate of $\eta(x_1,x_2,x_3)$ given that we have observed $D(a_1,a_2,a_3)$ is

$$\hat{\eta}(x_1,x_2,x_3|a_1,a_2,a_3) = y(x_1,a_2,a_3) + y(a_1,x_2,a_3) + y(a_1,a_2,x_3) - 2y(a_1,a_2,a_3) \qquad (5.3)$$

and since we observe true yields,

$$\hat{\eta}(x_1,x_2,x_3) = \eta(x_1,a_2,a_3) + \eta(a_1,x_2,a_3) + \eta(a_1,a_2,x_3) - 2\eta(a_1,a_2,a_3) .$$

It follows from above that the expression of the bias at

the point $(x_1,x_2,x_3)$ is given by

$$B(x_1,x_2,x_3|a_1,a_2,a_3) = \hat{\eta}(x_1,x_2,x_3|a_1,a_2,a_3) - \eta(x_1,x_2,x_3)$$

$$= \eta(x_1,a_2,a_3) + \eta(a_1,x_2,a_3)$$

$$+ \eta(a_1,a_2,x_3) - 2\eta(a_1,a_2,a_3)$$

$$- \eta(x_1,x_2,x_3) .$$

If we now express each of the true yields in terms of the identity 5.2, which is the true model, after cancellation we will have:

$$B(x_1,x_2,x_3|a_1,a_2,a_3) = \{[(E^{(1)}E^{(2)})_{x_1a_2} + (E^{(1)}E^{(2)})_{a_1x_2}$$

$$- (E^{(1)}E^{(2)})_{a_1a_2} - (E^{(1)}E^{(2)})_{x_1x_2}] + [(E^{(1)}E^{(3)})_{x_1a_3}$$

$$+ (E^{(1)}E^{(3)})_{a_1x_3} - (E^{(1)}E^{(3)})_{a_1a_3} - (E^{(1)}E^{(3)})_{x_1x_3}]$$

$$+ (E^{(2)}E^{(3)})_{x_2a_3} + (E^{(2)}E^{(3)})_{a_2x_3} - (E^{(2)}E^{(3)})_{a_2a_3}$$

$$- (E^{(2)}E^{(3)})_{x_2x_3}] + [(E^{(1)}E^{(2)}E^{(3)})_{x_1a_2a_3}$$

$$+ (E^{(1)}E^{(2)}E^{(3)})_{a_1x_2a_3} + (E^{(1)}E^{(2)}E^{(3)})_{a_1a_2x_3}$$

$$- 2(E^{(1)}E^{(2)}E^{(3)})_{a_1a_2a_3} - (E^{(1)}E^{(2)}E^{(3)})_{x_1x_2x_3}]\}.$$

We now square the above expression and average over all points $(x_1,x_2,x_3)$ of the lattice, and then average the resultant over all possible $D(a_1,a_2,a_3)$ stars by averaging over $(a_1,a_2,a_3)$. By using the relations

$$\sum_{x_1=1}^{L_1} E^{(1)}_{x_1} = \sum_{x_2=1}^{L_2} E^{(2)}_{x_2} = \ldots = \sum_{x_3=1}^{L_3} (E^{(2)}E^{(3)})_{x_2x_3}$$

$$= \ldots = \sum_{x_3=1}^{L_3} (E^{(1)}E^{(2)}E^{(3)})_{x_1x_2x_3} = 0,$$

it is found that

$$B = \frac{1}{L} \sum_{a_1,a_2,a_3}^{L_1,L_2,L_3} \sum_{x_1,x_2,x_3}^{L_1,L_2,L_3} \frac{B^2(x_1,x_2,x_3|a_1,a_2,a_3)}{L}$$

$$= \frac{L_1,L_3}{L} \sum_{x_1,x_2\ 1}^{L_1,L_2} (E^{(1)}E^{(2)})^2_{x_1,x_2} + \frac{4L_2}{L} \sum_{x_1,x_3\ 1}^{L_1,L_3} (E^{(1)}E^{(3)})^2_{x_1x_3}$$

$$+ \frac{4L_1}{L} \sum_{x_2,x_3\ 1}^{L_2,L_3} (E^{(2)}E^{(3)})^2_{x_2x_3}$$

$$+ \frac{8}{L} \sum_{x_1,x_2,x_3\ 1}^{L_1,L_2,L_3} (E^{(1)}E^{(2)}E^{(3)})^2_{x_1x_2x_3} \qquad (5.4)$$

We now consider the case in which each observation is equal to the true yield plus and additive error term, or

$$y(x_1,x_2,x_3) = \eta(x_1,x_2,x_3) + e(x_1,x_2,x_3),$$

and which the errors are assumed to be uncorrelated, with mean zero and the same variance $\sigma^2$.

According to 5.3

$$\hat{\eta}(x_1,x_2,x_3) = y(x_1,a_2,a_3) + y(a_1,x_2,a_3) + y(a_1,a_2,x_3)$$

$$- 2y(a_1,a_2,a_3),$$

so that

$$V[\hat{\eta}(x_1,x_2,x_3|a_1,a_2,a_3)] = 3\sigma^2 + 4\sigma^2$$

$$+ 2\delta(x_1,a_1)\delta(x_2,a_2)\sigma^2 + 2\delta(x_1,a_1)\delta(x_3,a_3)\sigma^2$$

$$+ 2\delta(x_2,a_2)\delta(x_3,a_3)\sigma^2 - 4\delta(x_1,a_1)\sigma^2 - 4\delta(x_2,a_2)\sigma^2$$

$$- 4\delta(x_3,a_3)\sigma^2$$

where $\delta(x_i,a_i)$ is the well-known Kronecker delta.

The average of the variance over all the points of the lattice given we estimate with the elements of the $D(a_1,a_2,a_3)$ star is

$$V_{a_1,a_2,a_3} = \frac{1}{L} \sum_{\substack{x_1,x_2,x_3 \\ 1}}^{L_1,L_2,L_3} V[\hat{\eta}(x_1,x_2,x_3|a_1,a_2,a_3)].$$

A small amount of computation yields

$$V_{a_1,a_2,a_3} = [7 + 2(\frac{L}{L_1L_2} + \frac{L}{L_1L_3} + \frac{L}{L_2L_3}) - 4(\frac{L}{L_1} + \frac{L}{L_2} + \frac{L}{L_3})]\sigma^2 \qquad (5.5)$$

Since the expression above is independent of the star chosen at random, the average variance over the set of all possible stars is this same expression.

From 5.4 and 5.5, the expression of the average mean square error is then:

$$J = [7 + 2(\frac{L}{L_1L_2} + \frac{L}{L_1L_3} + \frac{L}{L_2L_3}) - 4(\frac{L}{L_1} + \frac{L}{L_2} + \frac{L}{L_3})]\sigma^2$$

$$+ 4[\frac{L_3}{L} \sum_{\substack{x_1,x_2 \\ 1}}^{L_1,L_2} (E^{(1)}E^{(2)})^2_{x_1x_2} + \frac{L_2}{L} \sum_{\substack{x_1,x_3 \\ 1}}^{L_1,L_3} (E^{(1)}E^{(3)})^2_{x_1x_3}$$

$$+ \frac{L_1}{L} \sum_{\substack{x_2,x_3 \\ 1}}^{L_2,L_3} (E^{(2)}E^{(3)})^2_{x_2x_3}]$$

$$+ \frac{8}{L} [\sum_{\substack{x_1,x_2,x_3 \\ 1}}^{L_1,L_2,L_3} (E^{(1)}E^{(2)}E^{(3)})^2_{x_1x_2x_3}] \qquad (5.6)$$

3. The J criterion for pure sampling error

We now consider the general problem. We treat first the situation where there is just sampling error and so each observed value is equal to the true yield at the point.

Assuming we have $n$ factors $F^{(1)},F^{(2)},\ldots,F^{(n)}$ with $L_1,L_2,\ldots,L_n$ levels respectively, we can write the true yield at the lattice point $(x_1,x_2,\ldots,x_n)$ as

$$\eta(x_1,x_2,\ldots,x_n) = \mu + \sum_{r=1}^{n} E^{(r)}_{x_r} + \sum_{\substack{r<s \\ 1}}^{n} (E^{(r)}E^{(s)})_{x_r x_s}$$

$$+ \sum_{\substack{r<s<t \\ 1}}^{n} (E^{(r)}E^{(s)}E^{(t)})_{x_r x_s x_t} + \ldots +$$

$$+ \sum_{\substack{r_1<\ldots<r_k \\ 1}}^{n} (E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{x_{r_1}\ldots x_{r_k}} + \ldots +$$

$$+ (E^{(1)}E^{(2)}\ldots E^{(n)})_{x_1\ldots x_n} . \qquad (5.7)$$

We consider the case where the model assumed for the least squares fitting contains only main effects and it is given therefore by

$$y(x_1,x_2,\ldots,x_n) = \mu + E^{(1)}_{x_1} + E^{(2)}_{x_2} + \ldots + E^{(n)}_{x_n}$$

Notice that for $r = 1,2,\ldots,n$

$$(E^{(r)}_{x_r} - E^{(r)}_{a_r}) = (\mu + E^{(1)}_{a_1} + E^{(2)}_{a_2} + \ldots + E^{(r-1)}_{a_{r-1}} + E^{(r)}_{x_r}$$

$$+ E^{(r+1)}_{a_{r+1}} + \ldots + E^{(n)}_{a_n}) - (\mu + E^{(1)}_{a_1} + E^{(2)}_{a_2} + \ldots + E^{(r)}_{a_r}$$

$$+ \ldots + E^{(n)}_{a_n}) \ ,$$

so that the estimates of the above parametric functions are

$$(E^{(r)}_{x_r} - E^{(r)}_{a_r}) = y(a_1, a_2, \ldots, a_{r-1}, x_r, a_{r+1}, \ldots, a_n)$$

$$- y(a_1, a_2, \ldots, a_r, \ldots, a_n) \ .$$

Since the main effect model fitted is saturated, it follows that

$$\eta(x_1, x_2, \ldots, x_n) = (\mu + E^{(1)}_{a_1} + \ldots + E^{(n)}_{a_n} + (E^{(1)}_{x_1} - E^{(1)}_{a_1})$$

$$+ \ldots + (E^{(n)}_{x_n} - E^{(n)}_{a_n}) ,$$

and so from above

$$\hat{\eta}(x_1, x_2, \ldots, x_n | a_1, a_2, \ldots, a_n) = y(x_1, a_2, \ldots, a_n)$$

$$+ y(a_1, x_2, a_3, \ldots, a_n) + \ldots + y(a_1, a_2, \ldots, x_n)$$

$$- (n-1)\ y(a_1, a_2, \ldots, a_n) \ . \qquad (5.8)$$

Since in this first case we are assuming no error of observation

$$\eta(x_1,\ldots,x_n) = y(x_1,\ldots,x_n) ,$$

and from 5.8 we have

$$\hat{\eta}(x_1,\ldots,x_n|a_1,\ldots,a_n) = \eta(x_1,a_2,\ldots,a_n)$$

$$+ \eta(a_1,x_2,a_3,\ldots,a_n) +\ldots+ \eta(a_1,\ldots,a_{n-1},x_n)$$

$$- (n-1)\eta(a_1,\ldots,a_n) \qquad (5.9)$$

It is of some interest to note a property of the estimates of main effect with the star design and a main effect model. The estimate of the main effect of the factor $F^{(i)}$ at level $u_i$ is equal to

$$\hat{E}^{(i)}_{u_i} = y(a_1,\ldots,a_{i-1},u_i,a_{i+1},\ldots,a_n)$$

$$- \frac{1}{L_i} \sum_{v_i=1}^{L_i} y(a_1,\ldots,a_{i-1},v_i,a_{i+1},\ldots,a_n) .$$

This has an expectation given $(a_1,a_2,\ldots,a_n)$ equal to the same expression with y's replaced by $\eta$'s. The conditional expectation involves interactions of all degrees. We can see very quickly that the expectation of the estimate of any main effect, when expectation is taken over all vectors $(a_1,a_2,\ldots,a_n)$ is the main effect. This follows from the definition of a main effect which states that the main effect is exactly the average of the expressions above over all possibilities for $a_1,a_2,\ldots,a_{i-1},a_{i+1},\ldots,a_n$.

It follows from 5.9 that the bias at the point $(x_1,\ldots,x_n)$ given we observe the $D(a_1,\ldots,a_n)$ star is

$$B(x_1,\ldots,x_n|a_1,\ldots,a_n) \equiv \hat{\eta}(x_1,\ldots,x_n|a_1,\ldots,a_n)$$

$$- \eta(x_1,\ldots,x_n) = \eta(x_1,a_2,\ldots,a_n)$$

$$+\ldots+ \eta(a_1,\ldots,a_{n-1},x_n) - (n-1)\eta(a_1,\ldots,a_n)$$

$$- \eta(x_1,\ldots,x_n) .$$

Substituting for the true yields their corresponding expressions given by 5.7, we obtain after some reductions

$$B(x_1,x_2,\ldots,x_n|a_1,a_2,\ldots,a_n) = \{ \sum_{\substack{r<s \\ 1}}^{n} [(E^{(r)}E^{(s)})_{a_r a_s}$$

$$+ (E^{(r)}E^{(s)})_{x_r a_s} - (E^{(r)}E^{(s)})_{a_r x_s} - (E^{(r)}E^{(s)})_{x_r x_s}]$$

$$+ \sum_{\substack{r<s<t \\ 1}}^{n} [(E^{(r)}E^{(s)}E^{(t)})_{a_r a_s a_t} + (E^{(r)}E^{(s)}E^{(t)})_{a_r a_s a_t}$$

$$- 2(E^{(r)}E^{(s)}E^{(t)})_{a_r a_s a_t} - (E^{(r)}E^{(s)}E^{(t)})_{x_r x_s x_t}]$$

$$+\ldots+ \sum_{\substack{r_1<r_2<\ldots<r_k \\ 1}}^{n} [(E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{x_{r_1} a_{r_2}\ldots a_{r_k}}$$

$$+ (E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{a_{r_1} x_{r_2} a_{r_3}\ldots a_{r_k}}$$

$$+\ldots+ (E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{a_{r_1}\ldots a_{r_{k-1}}x_{r_k}}$$

$$- (k-1)(E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{a_{r_1}a_{r_2}\ldots a_{r_k}}$$

$$- (E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{x_{r_1}x_{r_2}\ldots x_{r_k}}]$$

$$+\ldots+ [(E^{(1)}E^{(2)}\ldots E^{(n)})_{x_1a_2\ldots a_n}$$

$$+ (E^{(1)}E^{(2)}\ldots E^{(n)})_{a_1x_2a_3\ldots a_n}$$

$$+\ldots+ (E^{(1)}E^{(2)}\ldots E^{(n)})_{a_1a_2\ldots a_{n-1}x_n}$$

$$- (n-1)(E^{(1)}E^{(2)}\ldots E^{(n)})_{a_1a_2\ldots a_n}$$

$$- (E^{(1)}E^{(2)}\ldots E^{(n)})_{x_1x_2\ldots x_n}]\} \qquad (5.10)$$

We now square the expression above and average over all the points $(x_1,\ldots,x_n)$ of the lattice to obtain the mean average square bias given $D(a_1,\ldots,a_n)$ and then average again over the points $(a_1,\ldots,a_n)$ to obtain the expected value of the average square bias over all possible stars, or

$$B = \frac{1}{L} \sum_{a_1,\ldots,a_n=1}^{L_1,\ldots,L_n} \sum_{x_1,\ldots,x_n=1}^{L_1,\ldots,L_n} \frac{B^2(x_1,\ldots,x_n|a_1,\ldots,a_n)}{L} .$$

We recall that

$$\sum_{x_i=1}^{L_i} (E^{(i)}E^{(j)})_{x_i x_j} = \sum_{x_j=1}^{L_j} (E^{(i)}E^{(j)})_{x_i x_j}$$

$$= \sum_{x_{r_i}=1}^{L_{r_i}} (E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{x_{r_1}\ldots x_{r_k}}$$

$$=\ldots= \sum_{x_n=1}^{L_n} (E^{(1)}E^{(2)}\ldots E^{(n)})_{x_1\ldots x_n} = 0. \qquad (5.11)$$

In the overall average of the square of the bias expression given by 5.10 we will obtain products of terms of different type. It is easy to see that when we consider the terms in two different brackets at least one pair $(x_i,a_i)$ occurs in the terms of one bracket and not in the other. It follows from 5.11 that summation of the bracket which contains $(x_i,a_i)$ over them alone gives zero. Hence the overall sum of the product of two different brackets is zero. Consider for example, the product

$$[(E^{(1)}E^{(2)})_{x_1a_2} + E^{(1)}_{a_1}E^{(2)}_{x_2} - (E^{(1)}E^{(2)})_{a_1a_2}$$

$$- (E^{(1)}E^{(2)})_{x_1x_2})] \times [(E^{(1)}E^{(3)})_{x_1a_3}$$

$$+ (E^{(1)}E^{(3)})_{a_1x_3} - (E^{(1)}E^{(3)})_{a_1a_3} - (E^{(1)}E^{(3)})_{x_1x_3}].$$

This is to be summed over all $(x_1,x_2,\ldots,x_n)$ and $(a_1,a_2,\ldots,a_n)$. If we consider summation first over $x_3$ and $a_3$, a pair which occurs in the second factor and not in the first, the second factor gives

$$\sum_{x_3,a_3}^{L_3} [(E^{(1)}E^{(3)})_{x_1a_3} + (E^{(1)}E^{(3)})_{a_1x_3}$$

$$- (E^{(1)}E^{(3)})_{a_1a_3} - (E^{(1)}E^{(3)})_{x_1x_3}]$$

and from the relations above, the sum of this term is zero.

We now consider the sum of the products of terms within a bracket. Notice that within a bracket two terms differ by at least one subscript and since the sum of any term on any single subscript is zero, it follows that the sum of the cross products of pairs of terms in the same bracket is zero. Hence now we are left only with the squares of the interactions inside each bracket. Consider the bracket involving k-th order interactions:

$$[(E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{a_{r_1}a_{r_2}\ldots a_{r_k}}$$

$$+ (E^{(r_1)}E^{(r_2)}\ldots E^{(r_n)})_{a_{r_1}x_{r_2}a_{r_3}\ldots a_{r_k}}$$

$$+\ldots+ (E^{(r_1)}E^{(r_2)}\ldots E^{(r_n)})_{a_{r_1}\ldots a_{r_{k-1}}x_{r_k}}$$

$$- (k-1)(E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{a_{r_1}a_{r_2}\ldots a_{r_k}}$$

$$- (E^{(r_1)}E^{(r_2)}\ldots E^{(r_k)})_{x_{r_1}x_{r_2}\ldots x_{r_k}}] \quad .$$

Each squared term involves the total sum of squares of the interactions of any particular type. The coefficient of the sum squared is clearly $k+(k-1)^2+1$, which equals $k^2-k+2$. The final expression for the average mean square bias is then

$$B = 4 \sum_{\substack{r<s \\ 1}}^{n} \frac{\prod_{j\neq r\neq s}^{n} L_j}{L} \sum_{\substack{x_r,x_s \\ 1}}^{L_r,L_s} (E^{(r)}E^{(s)})^2_{x_r x_s}$$

$$+ 8 \sum_{\substack{r<s<t \\ 1}}^{n} \frac{\prod_{j\neq r\neq s\neq t}^{n} L_j}{L} \sum_{\substack{x_r,x_s,x_t \\ 1}}^{L_r,L_s,L_t} (E^{(r)}E^{(s)}E^{(t)})^2_{x_r x_s x_t}$$

$$+\ldots+ (k^2-k+2) \sum_{r_1<\ldots<r_k}^{n} \frac{\prod_{j\neq r_1\neq\ldots\neq r_k}^{n} L_j}{L^{k-1}}$$

$$\sum_{x_{r_1},\ldots,x_{r_k}}^{L_{r_1},\ldots,L_{r_k}} (E^{(r_1)}\ldots E^{(r_k)})^2_{x_{r_1}\ldots x_{r_k}}$$

$$+ (n^2-n+2)\frac{1}{L^{n-1}} \sum_{x_1,\ldots,x_n}^{L_1,\ldots,L_n} (E^{(1)}\ldots E^{(n)})^2_{x_1\ldots x_n} \quad . \qquad (5.12)$$

The expression above can be written in a slightly simpler form as follows. Suppose a complete analysis of variance is made of the totality of numbers $\eta(x_1,x_2,\ldots,x_n)$. From this we shall obtain what we may term, $S_i$, $i=1,2,\ldots,n$ as the sum of squares due to the $i$-th order interactions. With this notation we have that

$$B = \frac{1}{L^{n-1}} \{4S_2 + 8S_3 +\ldots+ (k^3-k+2)S_k +\ldots+$$

$$+\ldots+ (n^2-n+2)S_n\}. \qquad (5.13)$$

4. <u>The J criterion in the case where there is sampling and observational errors</u>

We now consider the case where we have an additive random error associated with each observation, so that

$$y(x_1,x_2,\ldots,x_n) = \eta(x_1,x_2,\ldots,x_n) + e(x_1,x_2,\ldots,x_n),$$

where the random errors are assumed to be uncorrelated with mean zero and same variance $\sigma^2$. Since the average mean square bias is the same as in the absence of error, to obtain the general expression for the J-criterion in this case we will need only to add the average variance component which we will now derive.

According to 5.9 the expression of the estimated true yield at the lattice point $(x_1,x_2,\ldots,x_n)$ is given by

$$\hat{\eta}(x_1,x_2,\ldots,x_n|a_1,a_2,\ldots,a_n) = y(x_1,a_2,\ldots,a_n)$$

$$+ y(a_1,x_2,a_3,\ldots,a_n) +\ldots+ y(a_1,a_2,\ldots,a_{n-1},x_n)$$

$$- (n-1)y(a_1,a_2,\ldots,a_n).$$

This is a linear function of random variables with the same variance $\sigma^2$. The covariance of any two observations with different coordinates is zero. We can write immediately

$$\mathrm{Cov}[y(x_1,a_2,\ldots,a_n),\ y(a_1,x_2,a_3,\ldots,a_n)]$$

$$= \sigma^2\delta(x_1,a_1)\delta(x_2,a_2)$$

with other similar equations, where

$$\delta(u,v) = 1, \text{ if } u = v$$

$$= 0, \text{ otherwise.}$$

Similarly

$$\mathrm{Cov}[y(x_1,a_2,\ldots,a_n),\ y(a_1,a_2,\ldots,a_n)] = \sigma^2\delta(x_1,a_1)$$

with other similar equations. Hence

$$V[\hat{\eta}(x_1,x_2,\ldots,x_n)\,|\,(a_1,a_2,\ldots,a_n)]/\sigma^2 = n+(n-1)^2$$

$$+ 2\delta(x_1,a_1)\delta(x_2,a_2) + 2\delta(x_1,a_1)\delta(x_3,a_3)$$

$$+\ldots+ 2\delta(x_{n-1},a_{n-1})\delta(x_n,a_n) - 2(n-1)\delta(x_1,a_1)$$

$$- 2(n-1)\delta(x_2,a_2) -\ldots- 2(n-1)\delta(x_n,a_n)\ . \qquad (5.14)$$

We now obtain the expression of the average variance over all points of the lattice

$$V_{|a_1,a_2,\ldots,a_n} = \frac{1}{L} \sum_{\substack{x_1,x_2,\ldots,x_n \\ 1}}^{L_1,L_2,\ldots,L_n} V[\hat{\eta}(x_1,x_2,\ldots,x_n|a_1,a_2,\ldots,a_n)]$$

It follows from 5.13 above that

$$V_{|a_1,a_2,\ldots,a_n} = (n^2-n+1)\sigma^2 + \frac{1}{L} \sum_{\substack{s,s'=1 \\ s\neq s'}}^{n} \delta(x_s,a_s)\delta(x_{s'},a_{s'})\sigma^2$$

$$- \frac{2(n-1)}{L} \sum_{s=1}^{n} \sum_{\substack{x_1,\ldots,x_n \\ 1}}^{L_1,\ldots,L_n} \delta(x_s,a_s)\sigma^2\ .$$

Since the number of points with the same s-th and s'-th coordinates is $\prod_{\substack{j=1 \\ j\neq s\neq s'}}^{n} L_j$ and the number of points with the same s-th coordinate is $\prod_{\substack{j=1 \\ j\neq s}}^{n} L_j$, the expression for the variance reduces to

$$V_{|a_1,a_2,\ldots,a_n} = [n^2-n+1) + \sum_{\substack{r,s=1 \\ r\neq s}}^{n} \frac{L}{L_r L_s} - 2(n-1) \sum_{s=1}^{n} \frac{L}{L_s}]\sigma^2 \tag{5.15}$$

Notice that since $V_{|a_1,a_2,\ldots,a_n}$ is independent of the sampled point $(a_1,a_2,\ldots,a_n)$ it follows that its expected value over the population of stars is that given by 5.15.

Another general expression for this expected average variance seems worth presenting. We write down again the general expression of an estimated lattice point.

$$\hat{\eta}(x_1,x_2,\ldots,x_n|a_1,a_2,\ldots,a_n) = y(x_1,a_2,\ldots,a_n)$$

$$+ y(a_1,x_2,a_3,\ldots,a_n) +\ldots+ y(a_1,a_2,\ldots,a_{n-1},x_n)$$

$$- (n-1)y(a_1,a_2,\ldots,a_n) .$$

Notice that if in $\hat{\eta}(x_1,x_2,\ldots,x_n)$ we have $x_r = a_r$ for $r=1,\ldots,k$, the variance of the fitted value is

$$V[\hat{\eta}(x_1,x_2,\ldots,x_n|a_1,a_2,\ldots,a_n) = (n-k)\sigma^2 + (k-n+1)^2\sigma^2.$$

We can now partition the $L = L_1L_2\ldots L_n$ lattice points into disjoint sets of points with respectively $0,1,2,\ldots,n$ coordinates in common with the corresponding coordinates of $y(a_1,\ldots,a_n)$. It then follows that we will have $c_0 = \prod_{j=1}^{n}(L_j-1)$ points with no coordinates in common with $y(a_1,\ldots,a_n)$, $c_1 = \sum_{r=1}^{n}\prod_{\substack{j=1\\ j\neq r}}^{n}(L_j-1)$ points with one coordinate in common with $y(a_1,\ldots,a_n),\ldots$, $c_k = \sum_{r_1<\ldots<r_k}\prod_{\substack{j=1\\ j\neq r,\neq\ldots\neq r_k}}^{n}(L_j-1)$ points with $k$ coordinates in common with $y(a_1,\ldots,a_n)$ and so on, with $c_n=1$.

Of course,

$$c_0 + c_1 +\ldots+ c_n = L.$$

We can now write the variance function averaged over all points of the lattice as:

$$V_{|a_1,a_2,\ldots,a_n} = [(n^2-n+1)\sum_{k=0}^{n} c_k + \sum_{k=0}^{n} k(k-1)c_k - 2(n-1)\sum_{k=0}^{n} kc_k]\frac{\sigma^2}{L}. \qquad (5.16)$$

Since the above expression is independent of the star used for estimation, it follows that it is also the expression for the expected value of the average variance over the sampled

stars.

We now summarize the results above in the following theorem.

Theorem 5.1. Consider a factorial experiment with factors $F^{(1)},F^{(2)},\ldots,F^{(n)}$ with respectively $L_1,L_2,\ldots,L_n$ levels and the corresponding lattice of possible points of observation. If we choose a point $(a_1,a_2,\ldots,a_n)$ at random with equal probability assigned to every point and use the points of the star $D(a_1,a_2,\ldots,a_n)$ to estimate the remaining points of the lattice the expression of the expected value of the average mean square error over the population of star designs is given by

$$J = 4\sum_{\substack{r<s\\1}}^{n} \frac{\prod_{\substack{j\neq r\neq s\\1}}^{n} L_j}{L} \sum_{\substack{x_r,x_s\\1}}^{L_r,L_s} (E^{(r)}E^{(s)})^2_{x_r x_s}$$

$$+ 8\sum_{\substack{r<s<t\\1}}^{L_r,L_s,L_t} \frac{\prod_{\substack{j\neq r\neq t\\1}}^{n} L_j}{L} \sum_{\substack{x_r,x_s,x_t\\1}}^{L_r,L_s,L_t} (E^{(r)}E^{(s)}E^{(t)})^2_{x_r x_s x_t}$$

$$+ (k^2-k+2)\sum_{\substack{r_1<\ldots<r_k\\1}}^{n} \frac{\prod_{\substack{j\neq r_1\neq\ldots\neq r_k\\1}}^{n} L_j}{L} \sum_{\substack{x_{r_1},x_{r_2},\ldots,x_{r_k}\\1}}^{L_{t_1},L_{r_2},\ldots,L_{r_k}} (E^{(r_1)}\ldots E^{(r_k)})^2_{x_{r_1}\ldots x_{r_k}}$$

$$+\ldots+ (n^2-n+2)\frac{1}{L}\sum_{\substack{x_1,x_2,\ldots,x_n \\ 1}}^{L_1,\ldots,L_n} (E^{(1)}\ldots E^{(n)})^2_{x_1\ldots x_n}$$

$$+ [(n^2-n+1) + \sum_{\substack{r,s=1 \\ r\neq s}}^{n} \frac{L}{L_s L_r} + 2(n-1)\sum_{s=1}^{n}\frac{L}{L_s}]\sigma^2 . \qquad (5.17)$$

The portion of this involving $\sigma^2$ is the average variance and the portion not involving $\sigma^2$ is the mean average square bias.

Using the notation introduced in the last section we can also write the mean square error in the following form:

$$J = \frac{1}{L}[4S_2 + 8S_3 +\ldots+ (k^3-k+2)S_k +\ldots+$$

$$\ldots+ (n^2-n+2)S_n] +[(n^2-n+1) + \sum_{\substack{r,s=1 \\ r\neq s}}^{n} \frac{L}{L_s L_r} +$$

$$+ 2(n-1)\sum_{s=1}^{n}\frac{L}{L_s}]\sigma^2 .$$

# VI. RANDOM FRACTIONAL REPLICATION

## A. Introduction

### 1. Aims

We will now consider the case in which the lattice of points at which we take observations are the points of a standard fractional replication plan. The experimenter intending to perform a fractionally replicated experiment makes assumptions about the negligibility of certain effects and interactions. From the relationships which define effects and interactions in terms of the treatment combinations of the full factorial system and negligibility assumptions he obtains a reduced model involving effects and interactions which can be estimated with well defined subsets of treatment combinations, called fractions. A plausible general procedure is to choose at random one of the possible fractions of treatment combinations for use in the experiment. If the negligibility assumptions do not hold the estimates of the effects and interactions in the reduced model are biased. From the theory of least squares, we can obtain an expression for the bias in terms of the deleted effects and interactions, but general results of this nature are not available in the literature. Also, results which take into consideration the effect of a particular choice of a fraction on the nature of the bias have not been considered in the literature. In the present chapter we examine this spectrum of investigative

situations, and evaluate the J-criterion considered in the previous chapters for other situations.

## 2. General properties and definitions for the $p^n$ factorial system

Let p be a prime integer denoting the number of levels of each of the n factors under consideration, and let each treatment combination be represented by a n-tuple $x = (x_1,\ldots,x_n)$ where $x_i$, $i=1,\ldots,n$, is a positive integer modulo p, which denotes the level of the i-th factor. We then have $p^n$ n-tuples or treatment combinations describing the ways in which treatments and levels can be assigned to experimental units. We abstract the necessary general ideas from Kempthorne (1952) to which reference can be made for detailed exposition.

Factors are usually denoted by Latin capital letters such as A,B,C, and so on. The i-th level of the main effect A, denoted by $A_i$, is equal to the mean of true yields of all treatment combinations for which $x_1=i(p)$ minus the overall mean of the true yields which is denoted by $\mu$. Similarly the j-th level of an interaction $A^{\alpha_1}B^{\alpha_2}\ldots$, denoted by $(A^{\alpha_1}B^{\alpha_2}\ldots)_j$ is defined as the true mean of all treatment combinations for which $\sum_{i=1}^{n} \alpha_i x_i = j(p)$ minus $\mu$.

The linear form, $L(x) = \alpha_1 x_1 + \alpha_2 x_2 +\ldots+ \alpha_n x_n$, associated with each effect or interaction such as $A^{\alpha_1}B^{\alpha_2}\ldots$ splits the $p^n$ treatment combinations into p sets of $p^{n-1}$,

according to the value of $L(x)$ modulo $p$. There are in all $(p^n-1)/(p-1)$ effects and interactions, each with $(p-1)$ degrees of freedom. This number results from the fact that two $L(x)$'s differing only by a multiplicative constant induce the same partition.

The above definitions allow the expression of each treatment combination in terms of the mean, effects and interactions in the following way

$$\eta(x) = \mu + \sum_{j=0}^{N} \sum_{s=0}^{p-1} \delta_s^{L^{(j)}(x)} X_s^{(j)}$$

where $x = (x_1,\ldots,x_n)$, $M = (p^n-1)/(p-1)$, $X_s^{(j)}$ denotes the $s$-th level of the main effect or interaction $X^{(j)}$ and $L^{(j)}(x)$ the linear function corresponding to $X^{(j)}$, and $\delta_s^{L^{(j)}(x)} = 1$ if $L^{(j)}(x) = s$ and zero otherwise.

## 3. Fractional replications

Often we are restricted on the number of experimental units we can use and we cannot run a complete $p^n$ factorial.

In such cases, a fraction is chosen by observing only the combinations $x$ that simultaneously satisfy $k$ linearly independent equations

$$L^{(1)}(x) = u_1, \quad L^{(2)}(x) = u_2,\ldots,L^{(k)}(x) = u_k$$

where each $u_j \varepsilon \{0,1,2,\ldots,p-1\}$.

The linear forms $L^{(j)}(x)$, $j=1,2,\ldots,k$ must be chosen with care because the effects or interactions associated with these are then not estimable, or are confounded with the mean. Furthermore an effect or interaction corresponding to $L(x)$ is completely confounded with effects or interactions associated with

$$\nu[L(x) + \sum_{j=1}^{k} c_j L^{(j)}(x)]$$

for any permissible set of numbers $c_1, c_2, \ldots, c_k$ where $\nu$ is to be chosen so that the resultant linear form is one of the $(p^n-1)/(p-1)$ possible forms, which are conventionally chosen so that the first non-zero coefficient in the linear form is unity. It is this fact which necessitates care in the choice of $L^{(j)}(x)$, because it is usually desired to have no initial confounding of main effects or of these and two factor interactions.

We will refer to the $k$ linear independent effects or interactions which determine the fractions as the linear independent effects or interactions of the defining identity and we denote them by $X^{(1)}, \ldots, X^{(k)}$ and write

$$I = X^{(1)} = X^{(2)} = \ldots = X^{(k)} = Z^{(1)} = Z^{(2)} = \ldots$$

where $Z^{(1)}, Z^{(2)}, \ldots,$ consist of all the generalized interactions of $X^{(1)}$ to $X^{(k)}$. If $L^{(j)}(x)$, $j = 1,2,\ldots,k$ are the linear forms associated with $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$,

then the linear forms associated with the totality of X's in the defining identity are all the forms

$$\nu[c_1L_1(x) + c_2L_2(x) +\ldots+ c_kL_k(x)]$$

where each $c_i \varepsilon \{0,1,2,\ldots,p-1\}$ not all being zero and $\nu$ is the scalar multiplier necessary to make the first non-zero coefficient of an $x_i$ equal to unity. The total number of terms in this defining identity is $(p^k-1)/(p-1)$.

Since we have only $p^{n-k}$ treatment combinations in each fraction we can estimate only $(p^{n-k}-1)/(p-1)$ effects or interactions with them.

By considering the effects or interactions corresponding to all linear combinations of the linear forms of the linear independent of the effects or interactions of the defining identity and determining the effects or interactions corresponding to the sum of the linear forms of the effects and interactions of the reduced model with each of the linear forms mentioned above, the generalized interactions, we determine the confounding. In conformity with much of the literature a set of effects and interactions which are mutually confounded are called aliases of each other. Subsequently a model will be fitted picking only one or none of any set of mutually confounded effects or interactions.

We now give some general statements about the relationship between the partitions of the set of $p^n$ treatment

combinations determined by the effects and interactions.

Statement 1: The class of $p^{n-1}$ treatment combinations having the same level with respect to effect or interaction can be partitioned into $p$ disjoint sets of $p^{n-2}$ treatment combinations, each set having the same level with respect to any other effect or interaction. This is the well-known orthogonality property of prime power factors (Kempthorne, 1952).

Statement 2: Each effect or interaction which is not a member of the defining identity divides the $p^{n-k}$ treatment combinations of a fraction into $p$ disjoint sets of $p^{n-k-1}$ treatment combinations, each set having the same level with respect to the effect or interaction under consideration.

Proof: A $1/p^k$ fraction of a $p^n$ factorial is the set of $p^{n-k}$ treatment combinations $x = (x_1,\ldots,x_n)$ such that

$$L^{(1)}(x) = u_1(p),\ldots,L^{(k)}(x) = u_k(p) ,$$

where each $u_j$ is a member of the set $\{0,1,\ldots,p-1\}$, and $L^{(1)}(x),,,.L^{(k)}(x)$ denote the linear forms corresponding to the linear independent interactions of the defining identity. Every effect or interaction $X^{(j)}$ which is not a member of the defining identity is such that its corresponding linear form $L^{(j)}(x)$ cannot be written as a linear combination

of the linear forms of the elements of the defining identity. This implies that the system above and $L^{(j)}(x) = i_j(p)$ have $p^{n-k-1}$ solutions for every positive integer modulo $p$, and the statement above follows immediately.

Statement 3: Each alias of an effect or interaction $X$ partitions the treatment combinations of a fraction into the same sets as $X$.

Proof: Let us denote by $L^{(j_r)}(x)$, $r = 1,\ldots,p^k-1$ the linear forms corresponding to aliases of effect or interaction with linear form $L^{(j)}(x)$ which does not belong to the defining identity.

Notice now that $L^{(j_r)}(x)$ is given by

$$L^{(j_r)}(x) = \nu[L^{(j)}(x) + \sum_{p=1}^{k} c_\ell L^{(\ell)}(x)]$$

where $c_\ell$ is a modulo $p$ integer positive number and $L^{(\ell)}(x)$, $\ell = 1,\ldots,k$ are the linear independent linear forms corresponding to the linear independent elements of the defining identity, with the $c_\ell$'s not simultaneously equaling zero, and $\nu$ a scalar multiplier chosen so that the right-hand side is an admissible linear form.

The statement above will follow from the fact that $L^{(1)}(x),\ldots,L^{(k)}(x)$ assume constant values over all treatment combinations of a given fraction and so the set of treatment combinations on the fraction for which $L^{(j)}(x)$ and $L^{(j_r)}(x)$ are constant is the same.

Statement 4: In every fraction, at each level of an effect or interaction which does not belong to the defining identity, there are $p^{n-k-2}$ treatment combinations at each level of any other effect or interaction which also does not belong to the defining identity and is not an alias of the first effect or interaction.

Proof: Let us denote by $L^{(s)}(x)$ and $L^{(t)}(x)$, $s \neq t$, the linear functions corresponding to the two effects or interactions under consideration.

By Statement 2, $L^{(s)}(x)$ divides the $p^{n-k}$ treatment combinations of a fraction into $p$ disjoint classes of $p^{n-k-1}$ treatment combinations, where each class is determined by a system such as

$$L^{(1)}(x) = u_1(p), \ldots, L^{(k)}(x) = u_k(p)$$

and

$$L^{(s)}(x) = i_s(p)$$

with $u_1(p), \ldots, u_k(p)$ and $i_s(p)$ modulo $p$ positive integers.

Note now that $L^{(s)}(x)$ and $L^{(t)}(x)$ are not multiples of each other, i.e., $L^{(s)}(x) \neq i(p)\, L^{(t)}(x)$, and are not linear combinations of the linear forms of the elements of the defining identity.

It then follows that $L^{(t)}(x)$ divides the treatment combinations which satisfy the system above into $p$ disjoint

classes with $p^{n-k-2}$ treatment combinations each.

Statement 5: In each fraction an alias of an estimated effect or interaction of the reduced model has the same number of treatment combinations at each of its levels.

Proof: The linear function $L^{(j_r)}(x)$ corresponding to an alias of $L^{(j)}(x)$ can be written as

$$L^{(j_r)}(x) = \nu[L^{(j)}(x) + \sum_{e=1}^{k} c_e L^{(e)}(x)] .$$

Note that $\sum_{e=1}^{k} c_e L^{(e)}(x)$, $e = 1,\ldots,k$, are the linear forms corresponding to the linear independent effects or interactions of the defining identity. It then follows that $L^{(j_r)}(x)$ will assume $p$ different modulo $p$ positive integer values as $L^{(j)}(x)$ varies from $0$ to $p-1$.

## 3. Statement of the problem

In the present chapter we obtain the value of the J-criterion corresponding to fractional factorials in situations for which the interactions which are assumed negligible are not in fact so, and when the fraction used for fitting is chosen at random, with equal probability assigned to every fraction.

Once a particular fraction is chosen and the estimates of the parameters obtained, the J-criterion for that fraction will be evaluated by computing the average mean square error

resulting from prediction at the $p^n$ points of the lattice. We will then obtain the expected value of the J-criterion by averaging over the population of fractions.

4. Illustrative example

Suppose we are interested in a $3^3$ factorial system but can take only 9 observations. Assume also that we want to estimate only the main effects A,B,C and the interaction AC, and that the remaining two and three factor interactions are assumed to be zero.

In the situation described above it is immediate that a fractional replicate can be used to perform the experiment. Consider the 1/3 fraction of the totality of the $3^3$ treatment combinations in which the three factor interaction $ABC^2$ is confounded with the mean. The defining identity for this plan is given by

$$I = ABC^2$$

and the table of aliases is

$$\mu = ABC^2 = A^2B^2C$$
$$A = AB^2C = BC^2$$
$$B = AB^2C^2 = AC^2$$
$$C = AB = ABC$$
$$AC = AB^2 = BC.$$

It follows from the above, that the elements of each fraction are given by the points for which $L(x) = x_1+x_2+x_3 = u(3)$, where $u(3)$ stands for a modulo 3 positive integer. The elements of the fractions are then the following:

fraction number 0: (000), (011), (022), (101), (112), (120), (202), (210), (221);

fraction number 1: (100), (111), (122), (201), (212), (220), (010), (021), (002);

and

fraction number 2: (200), (211), (222), (001), (012), (020), (102), (110), (121),

where fractions number 0, 1 and 2 have the treatment combinations with respectively these levels with respect to $ABC^2$.

Without loss of generality let us consider the case where the fraction that is chosen for fitting is the first fraction.

The normal equations for the first fraction are given by

$$\begin{bmatrix} 9 & 3\theta' & 3\theta' & 3\theta' & 3\theta' \\ 3\theta & 3I & J & J & J \\ 3\theta & J & 3I & J & J \\ 3\theta & J & J & 3I & J \\ 3\theta & J & J & J & 3I \end{bmatrix} \begin{bmatrix} \mu \\ A_0 \\ A_1 \\ A_2 \\ B_0 \\ B_1 \\ B_2 \\ C_0 \\ C_1 \\ C_2 \\ AC_0 \\ AC_1 \\ AC_2 \end{bmatrix} = \begin{bmatrix} T \\ [A_0] \\ [A_1] \\ [A_2] \\ [B_0] \\ [B_1] \\ [B_2] \\ [C_0] \\ [C_1] \\ [C_2] \\ [AC_0] \\ [AC_1] \\ [AC_2] \end{bmatrix}$$

where $\theta' = (1,1,1)$, $J$ is a $3 \times 3$ matrix of unities, $I$ is a $3 \times 3$ identity matrix, and the square brackets enclosing a given effect or interaction level denote the total of all observations in the fraction with that level with respect to the effect or interaction under consideration.

By the definitions of the effects and interactions it follows that the sum of any effect or interaction over its levels is zero, so

$$\sum_{x_1=0}^{2} A_{x_1} = \sum_{x_2=0}^{2} B_{x_2} = \sum_{x_3=0}^{2} C_{x_3} = \sum_{x_1+x_3=0}^{2} AC_{x_1+x_3} = 0 .$$

Imposing these side conditions to the solutions of the

normal equations, we will have:

$$9\hat{\mu} = [T]$$

$$3\hat{\mu} + 3\hat{A}_0 = [A_0] \qquad 3\hat{\mu} + 3\hat{C}_0 = [C_0]$$

$$3\hat{\mu} + 3\hat{A}_1 = [A_1] \qquad 3\hat{\mu} + 3\hat{C}_1 = [C_1]$$

$$3\hat{\mu} + 3\hat{A}_2 = [A_2] \qquad 3\hat{\mu} + 3\hat{C}_2 = [C_2]$$

$$3\hat{\mu} + 3\hat{B}_0 = [B_0] \qquad 3\hat{\mu} + 3\hat{AC}_0 = [AC_0]$$

$$3\hat{\mu} + 3\hat{B}_1 = [B_1] \qquad 3\hat{\mu} + 3\hat{AC}_1 = [AC_1]$$

$$3\hat{\mu} + 3\hat{B}_2 = [B_2] \qquad 3\hat{\mu} + 3\hat{AC}_2 = [AC_2]$$

and it follows that the least squares estimates are given by

$$\hat{\mu} = [T]/9$$

$$\hat{A}_0 = [A_0]/3 - [T]/9 \qquad \hat{B}_0 = [B_0]/3 - [T]/9$$

$$\hat{A}_1 = [A_1]/3 - [T]/9 \qquad \hat{B}_1 = [B_1]/3 - [T]/9$$

$$\hat{A}_2 = [A_2]/3 - [T]/9 \qquad \hat{B}_2 = [B_2]/3 - [T]/9$$

$$\hat{C}_0 = [C_0]/3 - [T]/9 \qquad \hat{AC}_0 = [AC_0]/3 - [T]/9$$

$$\hat{C}_1 = [C_1]/3 - [T]/9 \qquad \hat{AC}_1 = [AC_1]/3 - [T]/9$$

$$\hat{C}_2 = [C_2]/3 - [T]/9 \qquad \hat{AC}_2 = [AC_2]/3 - [T]/9$$

Since in every fraction, in the set of treatment combinations of the same level with respect to an estimated effect or interaction we will find an equal number of treatment combinations at each level of another estimated effect or interaction, the form of the normal equations will be the same for each fraction. It follows that the form of the estimates of the effects and interactions is independent of the fraction.

If the effects and interactions not present in the reduced model are not negligible, the least squares estimators of the fitted parameters will be biased, and from the expressions above, it is easy to see that the expected values of the effects and interactions estimates will be as follows. Fitting with treatment combinations of the 0-th fraction we will have:

$$E(\hat{\mu}|0) = \mu + ABC_0^2$$

$$E(\hat{A}_0|0) = A_0 + AB^2C_0 + BC_0^2$$

$$E(\hat{A}_1|0) = A_1 + AB^2C_2 + BC_2^2$$

$$E(\hat{A}_2|0) = A_2 + AB^2C_1 + BC_1^2$$

$$E(\hat{B}_0|0) = B_0 + AB^2C_0^2 + AC_0^2$$

$$E(\hat{B}_1|0) = B_1 + AB^2C_1^2 + AC_2^2$$

$$E(\hat{B}_2|0) = B_2 + AB^2C_2^2 + AC_1^2$$

$$E(\hat{C}_0|0) = C_0 + AB_0 + ABC_0$$

$$E(\hat{C}_1|0) = C_1 + AB_1 + ABC_2$$

$$E(\hat{C}_2|0) = C_2 + AB_2 + ABC_1$$

$$E(\hat{AC}_0|0) = AC_0 + AB_0^2 + BC_0$$

$$E(\hat{AC}_1|0) = AC_1 + AB_2^2 + BC_2$$

$$E(\hat{AC}_2|0) = AC_2 + AB_1^2 + BC_1$$

The general form of the above equations is as follows. We let $u = 0,1,$ or $2$ denote the fraction that is used. We let $E(\hat{\theta}|u)$ denote the expectation of $\hat{\theta}$ as determined by the $u$-th fraction. Then inspection shows that for $u = 0, 1,2,$

$$E(\hat{\mu}|u) = \mu + ABC_u^2,$$

$$E(\hat{A}_{x_1}|u) = A_{x_1} + AB^2C_{2u+x_1} + BC^2_{u-x_1},$$

$$E(\hat{B}_{x_2}|u) = B_{x_2} + AB^2C^2_{u+x_2} + AC^2_{u-x_2},$$

$$E(\hat{C}_{x_3}|u) = C_{x_3} + AB_{u+x_3} + ABC_{u-x_3},$$

and (6.1)

$$E(\hat{AC}_{x_1+x_3}|u) = AC_{x_1+x_3}\, AB^2{}_{2(u+x_1+x_3)} + BC_{u-x_1-x_3}.$$

To indicate the general form that will be needed, we let $X^{(1)}, X^{(2)}, \ldots$ be the effects and interactions that are fitted. Denote by $(\hat{X}_i^{(j)} | u)$ the estimate of the i-th level of effect $X^{(j)}$ determined from the u-th fraction. Then the general structure is that

$$E(\hat{X}_i^{(j)} | u) = X_i^{(j)} + X_{i_1}^{(j_1)} + X_{i_2}^{(j_2)} ,$$

where $X^{(j_1)}$ and $X^{(j_2)}$ are the effects of interactions that are completely confounded with $X^{(j)}$ in the fractional plan, and the subscripts $i_1$ and $i_2$ are computable. The general form of relationships like the above will be given in Theorem 6.1.

The expectation of the estimator of a effect or interaction when we choose the fraction at random is given by:

$$E(\hat{X}_i^{(j)}) = E_u\{E(\hat{X}_i^{(j)} | u)\}$$

where $E(X_i^{(j)}) | u)$ stands for the expected value of the least squares estimator of $X_i^{(j)}$ given by the u-th fraction and $E_u$ for the expectation of the above over the population of fractions.

Let us now continue with our example and obtain the expectation over the population of fractions of the least squares estimators of the effects and interactions in the reduced model.

If for example we consider the effect A we have

$$E(\hat{A}_{x_1}) = \frac{1}{3} \sum_{u=0}^{2} (A_{x_1} + AB^2C_{2(u+x_1)} + BC^2_{u-x_1}) ,$$

and since

$$\sum_{\ell=0}^{2} AB^2C_\ell = \sum_{\ell=0}^{2} BC^2_\ell = 0$$

it follows that

$$E(\hat{A}_{x_1}) = A_{x_1}$$

so $\hat{A}_{x_1}$ is an unbiased estimator of $A_{x_1}$ with respect to observational error and random selection of a fraction.

It is easy to see that the above holds for any level of A as well as for any other estimator of effect or interaction of the reduced model, Theorem 6.2 proves this result in general.

We now obtain the expression of the average square bias resulting from fitting the reduced model when the complete model, i.e., the model with all effects and interactions included, is the true one.

The general form of the average square bias when using the estimates of parameters from the reduced model to predict at all $N=p^n$ points of the lattice by the $p^{n-k}$ treatment combinations from the u-th fraction, is given by

$$B_{(u)} = \frac{1}{N} \sum_{x=1}^{N} [E\hat{\eta}(x|u) - \eta(x)]^2$$

where $\eta(x)$ is the expected value of the treatment combination $x$ under the true model and $\hat{\eta}(x|u)$ is the predicted value at the same point based on the $u$-th fraction.

Let us now denote by $\Delta^{(u)}X_i^{(j)}$ the bias of the estimator of the $i$-th level of the interaction $X^{(j)}$ when we use the treatment combinations of the $u$-th fraction, $u = 0,1,2$.

We then note that the biases at individual points have the following nature.

For points of the 1-th fraction fitted by the observations from the 0-th fraction

$$E\hat{\eta}(x_1,x_2,x_3|0) - \eta(x_1,x_2,x_3) = (\Delta^{(0)}\mu - \Delta^{(1)}\mu) + (\Delta^{(0)}A_{x_1} - \Delta^{(1)}A_{x_1})$$
$$+ (\Delta^{(0)}B_{x_2} - \Delta^{(1)}B_{x_2}) + (\Delta^{(0)}C_{x_3} - \Delta^{(1)}C_{x_3}) + (\Delta^{(0)}AC_{x_1+x_3}$$
$$- \Delta^{(1)}AC_{x_1+x_3}) .$$

Similarly for points of the 2-th fraction the same expression with the superscript 1 in the $\Delta$'s replaced by 2.

In general for a point in the $v$-th fraction estimated from observations in the $u$-th fraction

$$E\hat{\eta}(x_1,x_2,x_3|u)-\eta(x_1,x_2,x_3)=(\Delta^{(u)}\mu-\Delta^{(v)}\mu)+(\Delta^{(u)}A_{x_1}-\Delta^{(v)}A_{x_1})$$

$$+(\Delta^{(u)}B_{x_2}-\Delta^{(v)}B_{x_2})+(\Delta^{(u)}C_{x_3}-\Delta^{(v)}C_{x_3})$$

$$+(\Delta^{(u)}AC_{x_1+x_3}-\Delta^{(v)}AC_{x_1+x_3}) \ .$$

Note that the sum of the biases of all levels of the estimators of an effect or interaction is zero. For the effect A, for example, we have

$$\Delta^{(u)}A_0+\Delta^{(u)}A_1+\Delta^{(u)}A_2 = \sum_{x_1=0}^{2}(ABC^2_{2(u+x_1)}+BC^2_{u-x_1}) = 0$$

If we denote summation over the treatment combinations of the v-th fraction by $\sum_{x\varepsilon\phi_v}$, it follows from Statement 1 and the remark about the bias given above, that the sum of the squared biases over the points of the v-th fraction when we estimate with the treatment combinations of the u-th fraction is given by:

$$S_v^{(u)} = \sum_{x\varepsilon\phi_v}[E\hat{\eta}(x|u)-\eta(x)] = 9[(\Delta^{(u)}\mu-\Delta^{(v)}\mu)^2]$$

$$+ 3[\sum_{x_1=0}^{2}(\Delta^{(u)}A_{x_1}-\Delta^{(v)}A_{x_1})^2 + \sum_{x_2=0}^{2}(\Delta^{(u)}B_{x_2}-\Delta^{(v)}B_{x_2})^2$$

$$+ \sum_{x_3=0}^{2}(\Delta^{(u)}C_{x_3}-\Delta^{(v)}C_{x_3}) +$$

$$+ \sum_{x_1+x_3=0}^{2} (\Delta^{(u)}AC_{x_1+x_3} - \Delta^{(v)}AC_{x_1+x_3})^2]$$

Hence, the average square bias resulting from fitting the reduced model by the treatment combinations of the u-th fraction and predicting at all $3^3$ points of this factorial system is

$$B_{(u)} = \frac{1}{27} \sum_{\substack{v=0 \\ v\neq u}}^{2} S_v^{(u)} ,$$

and the average over the fractions of this average square bias is

$$B = E_u\{B_{(u)}\} = \frac{1}{3} \sum_{u=0}^{2} B_{(u)} .$$

From the above

$$B = \frac{1}{9}[ \sum_{\substack{u,v=0 \\ u\neq v}}^{2} (\Delta^{(u)}{}_\mu - \Delta^{(v)}{}_\mu)^2]$$

$$\frac{1}{27}[ \sum_{x_1=0}^{2} \sum_{\substack{u,v=0 \\ u\neq v}}^{2} (\Delta^{(u)}A_{x_1} - \Delta^{(v)}A_{x_1})^2$$

$$+ \sum_{x_2=0}^{2} \sum_{\substack{u,v=0 \\ u\neq v}}^{2} (\Delta^{(u)}B_{x_2} - \Delta^{(v)}B_{x_2})^2$$

$$+ \sum_{x_3=0}^{2} \sum_{\substack{u,v=0 \\ u\neq v}}^{2} (\Delta^{(u)}C_{x_3} - \Delta^{(v)}C_{x_3})^2 +$$

$$+ \sum_{x_1+x_3=0}^{2} \sum_{\substack{u,v=0 \\ u\neq v}}^{2} (\Delta^{(u)}AC_{x_1+x_3} - \Delta^{(v)}AC_{x_1+x_3})^2]$$

Note now the general identity for a set of numbers $z_1, z_2, \ldots, z_n$:

$$\frac{1}{2N} \sum_{\substack{m,n=1 \\ m\neq n}}^{N} (z_m - z_n)^2 = \sum_{j=1}^{N} (z_j - \bar{z})^2, \quad \text{with} \quad \bar{z} = \frac{1}{N} \sum_{i=1}^{N} z_i \ .$$

As mentioned above, the biases of the estimates of an effect or interaction add to zero over the fractions. It follows that:

$$B = \frac{2}{3} \sum_{u=0}^{2} (\Delta^{(u)}\mu)^2 + \frac{2}{9} \sum_{\ell=0}^{2} \sum_{u=0}^{2} \{(\Delta^{(u)}A_\ell)^2 + (\Delta^{(u)}B_\ell)^2$$

$$+ (\Delta^{(u)}C_\ell)^2 + (\Delta^{(u)}AC_\ell)^2\} \ .$$

We now consider for example the expression for the sum of squares of the bias associated with A over its three levels:

$$\sum_{x_1=0}^{2} \sum_{u=0}^{2} (\Delta^{(u)}A_{x_1})^2 = \{(AB^2C_0 + BC_0^2)^2 + (AB^2C_2 + BC_1^2)^2$$

$$+ (AB^2C_1 + BC_2^2)^2 + (AB^2C_2 + BC_0^2)^2 + (AB^2C_1 + BC_0^2)^2$$

$$+ (AB^2C_0+BC_1^2)^2 + (AB^2C_1+BC_1^2)^2 + (AB^2C_0+BC_1^2)^2$$

$$+ (AB^2C_2+BC_2^2)^2\}.$$

From Statement 1 and because the sum of an effect or interaction over its level is zero, it follows that

$$\sum_{x_1=0}^{2}\sum_{u=0}^{2}(\Delta^{(u)}A_{x_1})^2 = 3\{(AB^2C_0)^2+(AB^2C_1)^2+(AB^2C_2)^2$$

$$+ (BC_0^2)^2+(BC_1^2)^2+(BC_2^2)^2\}.$$

We obtain similar expressions for the sum of squares of biases associated with its other effects and interactions that are fitted in the reduced model.

The sum of squares for the aliases of the mean is given by:

$$\sum_{u=0}^{2}(\Delta^{(u)}\mu)^2 = 18[(ABC_0^2)^2+(ABC_1^2)+(ABC_2^2)^2] .$$

Hence

$$B = \frac{2}{3}\sum_{\ell=0}^{2}\{(ABC_\ell^2)^2+(AB^2C_\ell)^2+(BC_\ell^2)^2+(AB^2C_\ell^2)^2$$

$$+ (AC_\ell^2)^2+(AB_\ell)^2+(ABC_\ell)^2+(AB_\ell)^2+(BC_\ell)^2\}.$$

Under the fitted reduced model the expected value of a fitted treatment combination is given by

$$\hat{\eta}(x|u) = \hat{\mu} + \hat{A}_{x_1} + \hat{B}_{x_2} + \hat{C}_{x_3} + AC_{x_1+x_3}$$

where, from the normal equations, the least squares estimates are given by:

$$\hat{A}_{x_1} = [A_{x_1}]/3 - [T]/9, \quad \hat{C}_{x_3} = [C_{x_3}]/3 - [T]/9$$

$$\hat{B}_{x_2} = [B_{x_2}]/3 - [T]/9, \quad AC_{x_1+x_3} = [AC_{x_1+x_3}]/3 - [T]/9.$$

It is easy to see that for any fraction

$$V[(\hat{\mu}|u] = \frac{\sigma^2}{9}, \quad \text{and } V[(\hat{A}_{x_1}|u)] = V[(\hat{B}_{x_2}|u)]$$

$$= V[(\hat{C}_{x_3}|u)] = [(AC_{x_1+x_3}|u)] = \frac{2\sigma^2}{9}$$

and that all these estimates are uncorrelated. Hence the variance of $\hat{\eta}(x|u)$ for any $x$ estimated from the $u$-th fraction is given by

$$V[\hat{\eta}(x)|u] = \frac{1}{9}\sigma^2 + 4 \times \frac{2}{9}\sigma^2 = \sigma^2 \quad .$$

Because this is the same for every point and for every fraction, the average of variance of prediction averaging over the points and the fractions is also $\sigma^2$. The general proof of this result is given in Theorem 6.5.

We can now write the final expression for the J-

criterion, when the fitted model uses the effects associated with A, B, C and AC as

$$J = \sigma^2 + \frac{2}{3} \sum_{\ell=0}^{2} [(ABC_\ell^2)^2 + (AB^2C_\ell)^2 + (BC_\ell^2)^2 + (AB^2C_\ell^2)^2 + (AC_\ell^2)^2 + (AB_\ell)^2 + (ABC_\ell)^2 + (AB_\ell^2)^2 + (BC_\ell)^2].$$

The average square bias is equal then to two-thirds of the sum of squares of true effects and interactions not included in the reduced model. This conclusions holds in general and is stated later as Theorem 6.3.

## B. General Expression for the J-Criterion for Randomized Fractional Replication of the $p^n$ System

### 1. Introduction

We give a general approach to the study of the properties of the estimators of effects and interactions as well as the value of the J-criterion for fractional replication of the $p^n$ factorial system when the reduced model is fitted and the complete model is the true one.

We note that the example treated in the preceding section is saturated, and we confine ourselves to this case in what

follows.

## 2. Expectation of the estimators of the parameters in the reduced model

The reduced model for the yield of a treatment combination is given by

$$y(x) = \mu + \sum_{\substack{e=1 \\ s=0,\ldots,(p-1)}}^{M} \sum_{s=0}^{p-1} \delta^{(e)}(x)\, X_s^{(e)} + e(x)$$

where

$$x = (x_1,\ldots,x_n), \quad M = (p^{n-k}-1)/(p-1) ,$$

$$X_s^{(e)}, \quad e = 1,\ldots,M, \quad \text{and} \quad \delta_s^{L^{(e)}}(x)$$

is as defined previously. Here $M$ is the number of effects and interactions in the fitted model and the $L^{(e)}(x)$ are their associated linear forms.

From the above we can write the design matrix for the treatment combinations of the $u$-th fraction, $u = 1,2,\ldots,p^k$, and it follows from Statements 1 through 4 that the normal equations for the estimation of the main effects and interactions of the reduced model are:

$$p^{n-k}\hat{\mu} + p^{n-k-1}\sum_{i=0}^{p-1}\hat{X}_i^{(1)} + \ldots + p^{n-k-1}\sum_{i}\hat{X}_i^{(M)} = [T]$$

$$p^{n-k-1}\hat{\mu} + p^{n-k-1}\hat{X}_0^{(1)} + p^{n-k-2}\sum_{i=0}^{p-1}X_i^{(2)} + \ldots + p^{n-k-2}\sum_{i=0}^{p-1}\hat{X}_1^{(M)}$$

$$= [X_0^{(1)}]$$

$$p^{n-k-1}\hat{\mu} + p^{n-k-1}\hat{X}_1^{(1)} + p^{n-k-2}\sum_{i=0}^{p-1}X_i^{(2)} + \ldots + p^{n-k-2}\sum_{i=0}^{p-1}X_i^{(M)}$$

$$= [X_1^{(1)}]$$

$$p^{n-k-1}\hat{\mu} + p^{n-k-1}\hat{X}_2^{(1)} + p^{n-k-2}\sum_{i=0}^{p-1}X_i^{(2)} + \ldots + p^{n-k-2}\sum_{i=0}^{p-1}\hat{X}_i^{(M)}$$

$$= [X_2^{(1)}]$$

$$\vdots$$

$$p^{n-k-1}\hat{\mu} + p^{n-k-2}\sum_{i=0}^{p-1}\hat{X}_i^{(1)} + \ldots + p^{n-k-2}\sum_{i=0}^{p-1}x_i^{(m-1)} + p^{n-k-1}\hat{X}_2^{(M)}$$

$$= [X_{p-1}^{(M)}]$$

We now impose as side conditions on the solutions the conditions coherent in the definitions of the effects and interactions, namely,

$$\sum_{i=0}^{p-1}\hat{X}_i^{(1)} = \sum_{i=0}^{p-1}\hat{X}_i^{(2)} = \ldots = \sum_{i=0}^{p-1}\hat{X}_i^{(M)} = 0 .$$

The normal equations then reduce to

$$p^{n-k}\hat{\mu} = [T]$$

$$p^{n-k-1}\hat{\mu} + p^{n-k-1}\hat{X}_0^{(1)} = [X_0^{(1)}]$$

$$p^{n-k-1}\hat{\mu} + p^{n-k-1}\hat{X}_1^{(1)} = [X_1^{(1)}]$$

$$p^{n-k-1}\hat{\mu} + p^{n-k-1}\hat{X}_2^{(1)} = [X_2^{(1)}]$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$p^{n-k-1}\hat{\mu} + p^{n-k-1}\hat{X}_{p-1}^{(M)} = [X_{p-1}]$$

and the solution is:

$$\hat{\mu} = [T]/p^{n-k}$$

$$\hat{X}_0^{(1)} = [X_0^{(1)}]/p^{n-k-1} - \hat{\mu}$$

$$\hat{X}_1^{(2)} = [X_1^{(1)}]/p^{n-k-1} - \hat{\mu}$$

$$\hat{X}_2^{(1)} = [X_2^{(1)}]/p^{n-k-1} - \hat{\mu}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\hat{X}_{p-1}^{(N)} = [X_{p-1}^{(M)}]/p^{n-k-1} - \hat{\mu}$$

Thus, the least squares estimator of the i-th level of the interaction $X^{(j)}$ is given by

$$\hat{X}_i^{(j)} = [X_i^{(j)}]/p^{n-k-1} - [T]/p^{n-k} .$$

Statement 4 says that the treatment combinations with level i with respect to any estimable $X^{(j)}$ are divided

into $p$ subsets of treatment combinations with different levels with respect to any other interaction which neither belongs to the defining identity nor is an alias of $X^{(j)}$. Also, for any effect or interaction $X^{(e)}$

$$\sum_{i=0}^{p-1} X_i^{(e)} = 0 .$$

It then follows from the above and the model expressions for the treatment combinations in terms of the effects and interactions in the true model that the expected value of the total of the observations with level $i$ with respect to $X^{(j)}$ will not have any contribution due to other estimated effects or interactions or their aliases.

From Statements 3 and 4, it follows that

$$E\left(\frac{[X_i^{(j)}]}{p^{n-k-1}}\right) = \mu + \sum_{e=1}^{E} X_{i_e}^{(\mu_e)} + X_i^{(j)} + \sum_{r=1}^{K} X_{i_r}^{(j_r)} ,$$

where

$$K = p^k - 1, \text{ and } R = \frac{p^k-1}{p-1} ,$$

$i_r$ is the level with respect to $X^{(j_r)}$ of any treatment combination with level $i$ with respect to $X^{(j)}$ in that fraction, and $X_{i_e}^{(\mu_e)}$, $e = 1,\ldots,R$, denote an effect or interaction which is completely confounded with the mean.

Similarly, by Statements 2, 3 and 4, it follows that

$$E\left(\frac{[T]}{p^{n-k}}\right) = \mu + \sum_{e=1}^{R} X_{i_e}^{(\mu_e)} \quad .$$

We then have:

$$E(\hat{X}_i^{(j)}|u) = X_i^{(j)} + \sum_{r=1}^{K} X_{i_r}^{(j_r)} = X_i^{(j)} + \Delta^{(u)}X_i^{(j)}$$

where $\Delta^{(u)}X_i^{(j)}$ stands for the biases of $X_i^{(j)}$ when we use the u-th fraction. Hence we have

Theorem 6.1: Suppose that a reduced model is fitted with the treatment combinations of the u-th $1/p^k$ fraction of a fractional replication and the complete model is the true one. Then the expected value of the estimate of $\mu$ is biased and given by

$$E(\hat{\mu}|u) = \mu + \Delta^{(u)}\mu, \quad \Delta^{(u)}\mu = \sum_{e=1}^{R} X_{i_e}^{(\mu_e)},$$

where $X^{(\mu_e)}$, $e = 1,\ldots,R = (p^k-1)/(p-1)$ are the aliases of $\mu$, and $i_e$ is the level of the treatment combinations of the u-th fraction with respect to $X^{(\mu_e)}$. The least squares estimator of the i-th level of any fitted effect or interaction is also biased, and its expected value is given by

$$E(\hat{X}_i^{(j)}|u) = X_i^{(j)} + \Delta^{(u)}X_i^{(j)}$$

where

$$\Delta^{(u)}X_i^{(j)} = \sum_{r=1}^{K} X_{i_r}^{(j_r)} , \quad K = p^k-1,$$

and $i_1,\ldots,i_k$ are the levels of the aliases $X^{(j_r)}$, $r = 1,\ldots,(p^k-1)$, associated with $X^{(j)}$ in the u-th fraction.

The linear form corresponding to an alias $X^{(j_r)}$ of $X^{(j)}$ is

$$L^{(j_r)}(x) = \nu[L^{(j)}(x) + \sum_{e=1}^{k} c_e L^{(e)}(x)] ,$$

where $L^{(e)}(x)$, $e = 1,\ldots,k$ are the linear forms corresponding to the effects and interactions of the defining identity. Thus, for $X_{i_r}^{(j)}$ we have

$$i_r = \nu[i + \sum_{e=1}^{k} c_e L^{(e)}(x)] = \nu[i+t(u)]$$

where $t(u)$ is constant for the u-th fraction and takes over the different fractions each of the values $0,1,\ldots,(p-1)$, $p^{k-1}$ times.

It then follows that $i_r$ assumes each of the values $0,1,\ldots,(p-1),p^{k-1}$ times, and since the sum of an effect or interaction over all its levels is 0, have the following result.

Statement 6: The sum over the fractions of the biases of a given level of an estimated effect or interaction is zero.

Our estimation procedure consists of first choosing a fraction at random with equal probability $1/p^k$ assigned to each fraction, and then fitting the reduced model using the treatment combinations of that fraction.

The general form of the expected value of the estimator of an effect or interaction is given by

$$E(\hat{X}_i^{(j)}) = E_u\{E(\hat{X}_i^{(j)} | u)\}$$

where $E(\hat{X}_i^{(j)} | u)$ stands for the expectation of the least squares estimator given that fraction $u$ was chosen and $E_u$ denotes expectation of over the fractions.

Using the general form of the expectation given above, from Theorem 6.1 and Statement 6, it follows that

$$E(\hat{X}_i^{(j)}) = \frac{1}{p^k} \sum_{u=1}^{k} (X_i^{(j)} + \Delta^{(u)} X_i^{(j)})$$

$$= X_i^{(j)} + \frac{1}{p^k} \sum_{u=1}^{p^k} \Delta^{(u)} X_i^{(j)} = X_i^{(j)} .$$

We can summarize the above as follows.

Theorem 6.2: If we randomly choose a fraction with equal probability, the estimator of the mean or any effect or interaction in the reduced model is unbiased with respect to observational error and random selection of a fraction.

3. The expectation of the average square bias

It follows from Theorem 6.1 that

$$E(\hat{x}_s^{(j)}|u) - x_s^{(j)} = \sum_{r=1}^{k} x_{i_s(r)}^{(j_r)} ,$$

where $K=p^k-1$, and $x^{(j_r)}$ for $r = 1,2,\ldots,k$ are the effects or interactions confounded with $x^{(j)}$ and $i_r(s)$ is the level of $x^{(j_r)}$ associated with level $s$ of $x^{(j)}$.

The linear function corresponding to any of the aliases of $x^{(j)}$ can be written as

$$L^{(j_r)}(x) = \nu[L^{(j)}(x) + \sum_{e=1}^{k} c_e L^{(e)}(x)]$$

where $L^{(e)}(x)$, $e = 1,\ldots,k$ are the linear independent interactions of the defining identity and its coefficients are positive integers modulo $p$.

Since $L^{(e)}(x)$, $e = 1,\ldots,k$ are constant for the elements of a particular fraction, we will have:

$$L^{(j_r)}(x) = \nu[L^{(j)}(x) + t(u)], \quad r = 1,\ldots,(p^k-1)$$

where $t(u)$ is the constant value of $\sum_{e=1}^{k} c_e L^{(e)}(x)$, corresponding to the $u$-th fraction.

It then follows from the above expression of $L^{(j_r)}(x)$ that for $L^{(j)}(x) = 0,1,\ldots,(p-1)$, $L^{(j_r)}(x)$ will take $p$ different values, which implies the following.

Statement 7: The sum of the biases of the estimates of the p levels of any fixed effect or interaction is zero, i.e.,

$$\sum_{i=0}^{p-1} \Delta^{(u)} X_i(j) = 0 .$$

The expression of the true yield in terms of the effects and interactions of the complete model is

$$\eta(x) = \mu + \sum_{m=1}^{P} \sum_{s=0}^{p-1} \delta_s^{L^{(m)}(x)} X_s^{(m)}, \quad \text{with} \quad P = (p^n-1)/(p-1),$$

where $X^{(m)}, m = 1,\ldots,(p^n-1)/(p-1)$ are the effects and interactions and the $L^{(m)}(x)$'s are their corresponding linear forms.

With the reduced model a predicted point is given by

$$\hat{\eta}(x|u) = \hat{\mu} + \sum_{e=1}^{M} \delta_s^{L^{(e)}(x)} \hat{X}_s^{(e)}, \quad M = (p^{n-k}-1)/(p-1) .$$

But since the complete model is the true one

$$E\hat{\eta}(x|u) = \mu + \Delta^{(u)}\mu + \sum_{e=1}^{M} \delta_s^{L^{(e)}(x)} (X_s^{(e)} + \Delta^{(u)} X_s^{(e)}) .$$

By Theorem 6.1 we can write the expected value of a fitted treatment combination in the v-th fraction under the complete model as

$$E\hat{\eta}(x|u) = \hat{\mu} + \Delta^{(v)}\mu + \sum_{e=1}^{M}\sum_{s=0}^{p-1}\delta_s^{L^{(e)}}(x)\,(X_s^{(e)}+\Delta^{(v)}X_s^{(e)})$$

where $M=(p^{n-k}-1)/(p-1)$ and the $X^{(e)}$'s are the effects and interactions of the reduced model.

The bias at a treatment combination of the v-th fraction after fitting the reduced model with the treatment combinations of the u-th fraction, is then given by

$$E\hat{\eta}(x|u) - \eta(x) = (\Delta^{(u)}\mu-\Delta^{(v)}\mu)$$

$$+ \sum_{e=1}^{M}\delta^{L^{(e)}}(x)\,(\Delta^{(u)}X_s^{(e)}-\Delta^{(v)}X_s^{(e)}) \ .$$

If we consider all treatment combinations of the v-th fraction, the sum of the square bias due to prediction is

$$S_v = \sum_{x\varepsilon\phi_v}[E\hat{\eta}(x|u)-\eta(x)]^2 = \sum_{x\varepsilon\phi_v}[(\Delta^{(u)}\mu-\Delta^{(v)}\mu)$$

$$+ \sum_{e=1}^{M}\delta_s^{L^{(e)}}(x)\,(\Delta^{(u)}X_s^{(e)}-\Delta^{(v)}X_s^{(e)}])^2$$

By Statements 4 and 5 we can write the above as:

$$S_v = p^{n-k}(\Delta^{(u)}\mu-\Delta^{(v)}\mu)^2+p^{n-k-1}\sum_{e=1}^{M}(\Delta^{(u)}X_0^{(e)}-\Delta^{(v)}X_0^{(e)})^2$$

$$+\ldots+ (\Delta^{(u)}X_{p-1}^{(e)}-\Delta^{(v)}X_{p-1}^{(e)})^2 \ .$$

We can then write the general form of the average square bias, resulting from fitting the reduced model with

the treatment combinations of the u-th fraction and predicting at all $p^n$ treatment combinations of the factorial system as:

$$B_{(u)} = \frac{1}{p^n} \sum_{\substack{v=1 \\ v\neq u}}^{p^k} S_v = \frac{1}{p^k} \sum_{\substack{v=1 \\ v\neq u}}^{p^k} (\Delta^{(u)}{}_\mu - \Delta^{(v)}{}_\mu)^2$$

$$+ \frac{1}{p^{k+1}} \sum_{\substack{v=1 \\ v\neq u}}^{p^k} \sum_{e=1}^{M} \sum_{i=0}^{p-1} (\Delta^{(u)} x_i^{(e)} - \Delta^{(v)} x_i^{(e)})^2$$

The expected value of the average square bias over the population of fractions is given by the expression:

$$B = E_u\{B_{(u)}\} = \frac{1}{p^k} \sum_{u=1}^{k} B_{(u)}$$

Substituting for $B_{(u)}$ in the above expression and using Statement 6 and the identity

$$\frac{1}{n} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} (y_i - y_j)^2 = 2 \sum_{i=1}^{n} (y_i - \bar{y})^2 ,$$

we obtain

$$B = \frac{2}{p^k} \sum_{u=1}^{p^k} (\Delta^{(u)}{}_\mu)^2 + \frac{2}{p^{k+1}} \sum_{u=1}^{p^k} \sum_{i=0}^{p-1} \sum_{j=1}^{M} (\Delta^{(u)} x_i^{(j)})^2$$

with $M = (p^{n-k}-1)/(p-1)$.

When we expand $\Delta^{(u)}{}_\mu$ and any of the terms of the type $\Delta^{(u)} x_i^{(j)}$ and square and then sum, because of the orthogonality

described in Statements 1 to 5, all cross product terms sum to zero and we find

$$B = \frac{2}{p}\left[\sum_{e=1}^{R}\sum_{i=0}^{p-1}(X_i^{(\mu_e)})^2 + \sum_{j=1}^{M}\sum_{i=0}^{p-1}\sum_{r=1}^{K}(X_i^{(j_r)})^2 2\right].$$

The above can be summarized as follows.

Theorem 6.3. Suppose a saturated reduced model is fitted from one observation on each treatment combination of a $1/p^k$ fraction of a fractional replication plan chosen at random with probability $1/p^k$, then the average square bias over the full set of lattice points and the population of fractions is given by

$$B = \frac{2}{p}\,[\text{sum of squares of effects and interactions ignored in the fitted model}].$$

4. The expectation of the average variance

We next obtain the average variance component of the J-criterion. Using the same notation as earlier, under the reduced model we can write a predicted treatment combination value as:

$$\hat{\eta}(x|u) = \hat{\mu} + \sum_{e=1}^{M} \hat{X}_{i_e}^{(e)} .$$

By the normal equations, the least squares estimators of $\mu$ and $X_{ij}^{(j)}$ are respectively given by

$$\hat{\mu} = [T]/p^{n-k} \quad \text{and} \quad \hat{x}_{ij}^{(j)} = [x_{ij}^{(j)}]/p^{n-k-1} - [T]/p^{n-k},$$

It easily follows that

$$V[\hat{\mu}] = \frac{\sigma^2}{p^{n-k}} \quad \text{and} \quad V[\hat{x}_{ij}^{(j)}] = \frac{p-1}{p^{n-k}}\ \sigma^2 \ .$$

Further, from Statement 1 it can also be seen that

$$\mathrm{Cov}[\hat{\mu},\hat{x}_{i_j}^{(j)}] = 0 \quad \text{and} \quad \mathrm{Cov}[\hat{x}_{i_j}^{(j)},\hat{x}_{i_{j'}}^{(j')}] = 0$$

for all $j \neq j'$.

Thus, since there are $(p^{n-k}-1)/(p-1)$ effects and interactions in the fitted model, it follows that

$$V[\hat{\eta}(x|u)] = \frac{\sigma^2}{p^{n-k}} + \left(\frac{p^{n-k}-1}{p-1}\right)\left(\frac{p-1}{p^{n-k}}\right)\sigma^2 = \sigma^2 \ .$$

It is immediate from the above that the average variance in fitting all $p^n$ treatment combinations with the effects and interactions estimates obtained from the $u$-th fraction is given by:

$$V_{(u)} = \frac{1}{p^n} \sum_{i=1}^{p^n} V[\hat{\eta}(x|u)] = \sigma^2$$

It is then also immediate that if we choose the fractional replicate at random as described earlier the expression of the expected value of the average variance over the population of fractions is

$$V = \sigma^2 \ .$$

From the result above and Theorem 6.5 we can state the following final result.

Theorem 6.3. Suppose a saturated reduced model is fitted from one observation on each treatment combination of a $1/p^k$ fraction of a fractional replication plan chosen at random with probability $1/p^k$, then the mean square error over the full set of lattice points and the population of fractions is given by

$$J = \sigma^2 + \frac{2}{p} \text{ [sum of squares of effects and interactions ignored in the fitted model].}$$

Theorem 6.4 holds for the case in which the effects and interactions fitted to a fraction give a saturated model. To extend the result to the case of fitting an unsaturated model, the following reasoning is useful. Suppose the unsaturated model that is fitted contains the effects and interactions $X^{(j)}$, $j = 1,2,\ldots,M'$, with $M' < M = (p^{n-k}-1)/(p-1)$. Then we can complete the fitted set to a saturated set by adjoining effects and interactions $X^{(j)}$, $j = M'+1, M'+2, \ldots, M$.

We then have

$$E\hat{\eta}(x|u) = E(\hat{\mu}|u) + \sum_{j=1}^{M'} E(\hat{X}_i^{(j)})$$

Also for $x$ belonging to the $v$-th fraction

$$\eta(x) = \mu(v) + \sum_{m=1}^{P} \sum_{s=0}^{p-1} \delta_s L^{(m)}(x) X_s^{(m)}(v)$$

where

$$X_s^{(j)}(v) = E(\hat{X}_s^{(j)} | v).$$

We have seen that

$$E(\hat{\mu}|u) = \mu + \Delta^{(u)}\mu,$$

and

$$E(\hat{X}_{i_j}^{(j)}|u) = X_s^{(j)} + \Delta^{(u)}X_{\cdot s}^{(j)} .$$

Hence

$$E\hat{\eta}(x|u)-\eta(x) = (\Delta^{(u)}\mu-\Delta^{(v)}\mu)$$

$$+ \sum_{e=1}^{M'} \sum_{s=0}^{p-1} \delta_s^{L^{(e)}(x)} [\Delta^{(u)}X_s^{(e)}-\Delta^{(v)}X^{(e)}{}_s]$$

$$- \sum_{e=M'+1}^{M} \sum_{s=0}^{p-1} \delta_s^{L^{(e)}(x)} [X_s^{(e)}+\Delta^{(v)}X_s^{(e)}] .$$

This expression must be squared, averaged over all x's and over all u. As before all crossproducts of terms from different effects sum to zero, because of the orthogonality properties.

The contributions from the effects and interactions that are fitted are the same as in the saturated case. For the non-fitted effects and interactions which appear with negative sign in the last equality, we use the following derivation.

We use $\sum_{x\varepsilon\phi_v}$ to indicate summation over the x's that

are in the v-th fraction. Then the contribution to the overall sum from the negative term in the last expression is

$$\sum_{u=1}^{p^k} \sum_{v=1}^{p^k} \sum_{x\varepsilon\phi_v} \{ \sum_{s=0}^{p-1} \delta_s^{L^{(e)}(x)} [X_s^{(e)} + \Delta^{(u)} X_s^{(e)}] \}^2$$

$$= \sum_{u=1}^{p^k} \sum_{v=1}^{p^k} \sum_{x\varepsilon\phi_v} \{ \sum_{s,s'=0}^{p-1} \delta_s^{L^{(e)}(x)} \delta_{s'}^{L^{(e)}(x)} [X_s^{(e)} + \Delta^{(v)} X_s^{(e)}]$$

$$[X_{s'}^{(e)} + \Delta^{(v)} X_{s'}^{(e)}]$$

$$= \sum_{u=1}^{p^k} \sum_{v=1}^{p^k} \sum_{s,s'=0}^{p-1} [X_s^{(e)} + \Delta^{(v)} X_s^{(e)}]$$

$$[X_{s'}^{(e)} + \Delta X_{s'}^{(e)}] \sum_{x\varepsilon\phi_v} \delta_s^{L^{(e)}(x)} \delta_{s'}^{L^{(e)}(x)} .$$

But

$$L_s^{(e)}(x) \; \delta_{s'}^{L^{(e)}(x)} = \delta_s^{L^{(e)}(x)} \; \delta_{ss'},$$

because a particular $x$ is associated with one of the levels of any effect or interaction.

This term gives therefore

$$\sum_{u=1}^{p^k} \sum_{v=1}^{p^k} \sum_{s=0}^{p-1} \sum_{x\varepsilon\phi_v} [X_s^{(e)} + \Delta^{(v)} X_s^{(e)}]^2 \; \delta_s^{L^{(e)}(x)}$$

$$= p^k \sum_{v=1}^{p^k} \sum_{s=0}^{p-1} \sum_{x\varepsilon\phi_v} [X_s^{(e)} + \Delta^{(e)} X_s^{(e)}]^2 \delta_s^{L^{(e)}(x)}$$

$$= p^k \sum_{s=0}^{p-1} \sum_{v=1}^{p^k} \delta_s^{L^{(e)}}(x) \left[ (X_s^{(e)})^2 + 2(X_s^{(e)})(\Delta^{(v)} X_s^{(e)}) \right.$$

$$\left. + (\Delta^{(v)} X_s^{(e)})^2 \right] \qquad (6.2)$$

The summation involving crossproducts may be written as

$$2 \sum_{s=0}^{p-1} X_s^{(e)} \sum_{v=1}^{p^k} \sum_{x \varepsilon \phi_v} \delta_s^{L^{(e)}}(x) \Delta^{(v)} X_s^{(e)}$$

$$= 2 \sum_{s=0}^{p-1} X_s^{(e)} \sum_{v=1}^{p-1} \Delta^{(v)} X_s^{(e)} \; p^{n-2} = 0$$

because by Statement 5

$$\sum_{v=1}^{p^k} \Delta^{(v)} X_s^{(e)} = 0,$$

for any $X^{(e)}$ of a saturated model. Similarly the last term of 6.2 gives

$$p^k \sum_{s=0}^{p-1} \sum_{v=1}^{p^k} \sum_{x \varepsilon \phi_v}' \delta_s^{L^{(e)}}(x) \left[ \sum_{r=1}^{K} X_{i_r}^{(j_r)} \right]^2 ,$$

where $K = p^k - 1$.

A similar expression and rearrangement shows that all cross-product terms here sum to zero. Consequently, the square of the negative term in 6.2 gives

$p^{n+k-1}$ [sum of squares of non-fitted effects and interactions of a saturated model and their aliases],

and we have to divide by $p^{n+k}$. Consequently, the B measure is equal to

$\frac{2}{p}$ [sum of squares of effects and interactions that are aliased with the mean or fitted effects and interactions]

$+ \frac{1}{p}$ [sum of squares of all effects and interactions which are not fitted or are not aliased with the mean and fitted effects and interactions].

Since the estimates of $\mu$ and any fitted effect or interaction are the same as in the saturated case, and these estimates are uncorrelated it follows that when we fit $M' < (p^{n-k}-1)/(p-1)$ effects or interactions the variance of a fitted point is given by

$$V[\hat{\eta}(x|u)] = \frac{\sigma^2}{p^{n-k}} + \frac{M'(p-1)}{p^{n-k}} \sigma^2 ,$$

and this will also be the value for the expected average variance because it does not depend on $u$.

We can summarize the results above on bias and variance as follows.

Theorem 6.5. Suppose a reduced model involving $M'$ effects and interactions is fitted from one observation on each treatment combination of a $1/p^k$ fraction of a fractional replication plan chosen at random with probability $1/p^k$, then the mean square error over the full set of lattice points and the population of fractions is given by

$$J = \frac{[1+M'(p-1)]}{p^{n-k}} \sigma^2 + \frac{2}{p} S_I + \frac{1}{p} S_{II}$$

where

$S_I$ = the sum of squares of effects and interactions that are confounded with the mean or fitted effects and interactions, and

$S_{II}$ = the sum of squares of all effects and interactions that are not fitted or not confounded with the mean or fitted effects and interactions.

We note that if the fitted model is saturated, then $M'=(p^{n-k}-p)/(p-1)$, the variance contribution becomes $\sigma^2$, the term $S_{II}$ becomes null, and we have the result of Theorem 6.4. It is easy to see that the argumentation used above for the case of $p$ levels with $p$ prime number can be modified directly for the case of $p^m$ levels of each factor with $p$ prime and $m$ an integer greater than unity. The only difference is that the linear forms are defined over the

galois field of $p^m$ elements rather than the galois field of $p$ elements.

It is of some interest to compare the result for the saturated case with the corresponding result for the random star-design described in Chapter V which is also saturated. It is immediate that the average variance of the fractional design is smaller than the variance of the star design. To compare the biases of the two designs we note that the sum of squares due to interactions of $r$-th order in the analysis of variance of the whole set of $\eta(x)$'s, which was denoted by $S_r$ is equal to $p^{n-1}S(r)$, where $S(r)$ is the sum of squares of the factorially defined corresponding interactions.

Hence for the star design

$$B = \frac{1}{p}[4S(2) + 8S(3) +\ldots+ (r^2-r+2)S(r)+\ldots+(n^2-n+2)S(n)]$$

while for the fractional design

$$B = \frac{2}{p}[S(2)+S(3) +\ldots+ S(r) +\ldots+ S(n)] .$$

Clearly the star design has an average square bias which is at least twice the average square bias of the fractional design.

# VII. SUMMARY AND CONCLUSIONS

Model fitting consists in the determination of the best relation between the independent variables $x_1, x_2, \ldots, x_n$ and a dependent response $y$. Usually two important kinds of errors arise in this process: error due to bias and error due to variance. By bias error is meant the error that is due to wrong model assumptions, when the mathematical function which is chosen to express the relation between the $x$'s and $y$ is not the true one. If instead of observing the true response value we observe it with error having some random structure we are said to have variance error. If the proper model is fitted and the errors have expectation zero, the method of least squares provides parameter estimates that on the average are equal to the true values. If the proper model is not fitted this is not so in general, and the model fitted will be biased. We can, however, reduce the possible biases by choice of pattern of observation, that is, by choice of the design.

A major part of this dissertation was directed to the choice of design in relation to criteria of value of the fitted model. Several functions of the bias or of criteria related to bias can be defined. The first we focused our attention on was the maximum bias criterion, which gives the maximal deviation between the fitted and the true model over the region in factor space of interest. The case considered

was the one in which there is just one independent variable which lies inside a finite interval. Furthermore we restrict ourselves only to the fitting of polynomial models and the problem was how to allocate the observations to minimize the maximal bias when the fitted polynomial has degree s and the true polynomial has degree t>s. Results were obtained for the case when the polynomial fitted has degree s and the true polynomial has degree s+1. An optimal design for this criterion which received special attention consisted of the allocation which places equal number of observations at the zeros of a Chebyshev polynomial of degree equal or greater than s+1. Besides being optimal for the maximum bias criterion these designs have the important feature of providing good polynomial approximation even when the true underlying model is an unknown continuous function having at least two derivatives at each point of the interval. If we use the Chebyshev allocation and a linear combination of Chebyshev polynomials is fitted to the data these designs provide orthogonality or independent estimates of every parameter in the fitted model. It is also shown that there does not exist an optimal allocation for the maximum bias criterion when the difference between the degree of the fitted model and the true model is greater than one.

The next problem considered is related to the fitting of a polynomial model of degree lower than the one of the

true underlying polynomial model. The criterion under consideration was the average square bias considered by Box and Draper (1959) and Folks (1958). For the designs obtained a linear function of the sum of squares due to lack of fit can be used to estimate the average square bias and to give the experimenter an idea of how far off on the average his fitted model is from an assumed true model. These designs also minimize the average square bias criterion. A detailed study is made of the optimal designs with respect to the above criterion and it is shown that for the one-dimensional case these designs have $(\text{bias})^2$ and variance functions which have the undesirable feature of assuming very large values at the extremal portions of the interval. This motivates the consideration of criteria which correct that deficiency by giving more weight to those portions of the interval. Conditions are obtained for optimal designs with respect to the average square bias weighted with a positive function with respect to which all moments exist. In particular, it is shown that if three well-known classes of probability density functions of the beta, the gamma, and the normal are taken as weight functions the designs which minimize the average weighted bias when a polynomial of degree $s$ is fitted and the true model has degree $s+1$ place the observations at the zeros of Jacobi, Laguerre, and Hermite polynomials of degree $s+1$ respectively. A special case of the

above is the beta Be(1/2,1/2) density function. It is shown that the allocation at the zeros of a Chebyshev polynomial of degree $n \geq s+t$ is also optimal with respect to the average square bias weighted with the Chebyshev weight when the model fitted has degree $s$ and the true model has degree $t$, with $t>s$. It is also proved that optimal designs with respect to this criterion and the maximum bias designs are the same. The considerations with respect to polynomial models are concluded with comparisons between the optimal designs obtained in this study and other designs available in the literature. The comparisons are made not only with respect to several criteria related to bias but also with respect to criteria related to variance. The general conclusion is that one-dimensional designs which are optimal with respect to criteria related to variance give poor efficiency with respect to criteria related to bias, while optimal designs for criteria related to bias have at least acceptable performance with respect to variance. It is also shown that the optimal designs with respect to the maximum bias criterion or minimax bias designs are highly efficient when both bias and variance errors are present and so are the ones indicated for use in practical applications.

We next considered the case where the dependent variable is a function of the levels of qualitative factors and a fraction of the totality of points in the factorial

system is chosen for estimation and testing purposes. In order to obtain useful average properties, an element of randomization is introduced in the design. Random star designs which permit the estimation of all main effects are introduced and a general form for the mean square error resulting from the prediction at all points of the factorial space is obtained.

An alternative design possible for prime power factorial systems is the use of a random fraction defined by finite geometry considerations. Such designs are called random fractional replication designs. Biases and variances are obtained for this class of designs. The overall mean square error was found to have a very simple form, in terms of variance of observation and sums of effects or interactions that are ignored in the fitted model. It is noted that for the case when both star and random fractional designs are saturated, the latter is clearly superior.

# VIII. LITERATURE CITED

Abramovitz, A. and Stegun, I. A. 1968. Handbook of mathematical functions. New York, N.Y., Dover Publications, Inc.

Addelman, S. 1961. Irregular fractions of the $2^n$ factorial experiments. Technometrics 3: 479-496.

Addelman, S. 1962. Orthogonal main-effect plans for asymmetrical factorial experiments. Technometrics 4: 26-41.

Addelman, S. and Kempthorne, O. 1961. Some main-effect plans and orthogonal arrays of strength two. Annals of Mathematical Statistics 32: 1167-1176.

Banerjee, K. S. 1950. A note on fractional replication of factorial arrangements. Sankya 10: 87-94.

Bose, R. C. 1947. Mathematical theory of symmetrical factorial design. Sankya 8: 107-166.

Bose, R. C. and Bush, K. A. 1952. Orthogonal arrays of strength two and three. Annals of Mathematical Statistics 23: 508-524.

Bose, R. C. and Kishen, K. 1940. On the problem of confounding in the general symmetrical factorial design. Sankya 5: 21-36.

Box, G. E. P. and Draper, N. R. 1959. A basis for selection of a response surface design. Journal of the American Statistical Association 54: 622-654.

Box, G. E. P. and Hunter, T. S. 1961a: The $2^{k-p}$ fractional factorial designs I. Technometrics 3: 311-352.

Box, G. E. P. and Hunter, T. S. 1961 : The $2^{k-p}$ fractional factorial designs II. Technometrics 3: 449-458.

Chakravarti, I. M. 1956. Fractional replication in asymmetrical factorial designs and partially balanced arrays. Sankya 17: 143-164.

Daniel, C. 1956. Fractional replication in industrial research. Proceedings of the third Berkeley symposium on mathematical statistics and probability volume V: 87-98. Berkeley, California, Universities of California Press.

De La Garza, A. 1954. Spacing of information in polynomial regression. Annals of Mathematical Statistics 29: 294-298.

Dempster, A. P. 1960. Random allocation designs I: On general classes of estimation methods. Annals of Mathematical Statistics 31: 885-905.

Dempster, A. P. 1961. Random allocation designs II: Approximate theory for simple random allocation. Annals of Mathematical Statistics 32: 387-405.

Dystra, O. 1959. Partial duplication of factorial experiments. Technometrics 1: 63-75.

Ehrenfeld, S. 1955. On the efficiency of experimental designs. Annals of Mathematical Statistics 26: 247-297.

Ehrenfeld, S. and Zacks, S. 1961. Randomization and fractional experiments. Annals of Mathematical Statistics 32: 270-297.

Ehrenfeld, S. and Zacks, S. 1963. Optimal strategies in factorial experiments. Annals of Mathematical Statistics, 34: 780-791.

Ehrenfeld, S. and Zacks, S. 1967. Testing hypothesis in randomized factorial experiments. Annals of Mathematical Statistics 30: 1494-1507.

Elfving, G. 1952. Optimum allocation in linear regression theory. Annals of Mathematical Statistics 23: 255-262.

Farrel, R. H. 1968. On the admissibility at $\infty$, within the class of randomized designs, of balance designs. Annals of Mathematical Statistics 39: 1978-1994.

Farrel, R. H. 1969. On the admissibility of a randomized symmetrical design for the problem of a one way classification. Annals of Mathematical Statistics 40: 356-365.

Farrel, R. H. and Kiefer, J. and Walbran, A. 1965. Optimum multivariate designs. Fifth Berkely Symposium on Mathematical Statistics and Probability, Volume 1: 83-112.

Finney, D. J. 1945. The fractional replication of factorial experiments. Annals of Eugenics 12: 291-301.

Fisher, R. A. 1943. Theory of confounding in factorial experiments in relation to theory of groups. Annals of Eugenics 11: 341-353.

Fisher, R. A. 1945. A system of confounding for factors with more than two alternatives giving completely orthogonal cubes and higher powers. Annals of Eugenics 12: 283-290.

Folks, J. L. 1958. Comparison of designs for exploration of response relationships. Unpublished Ph.D. thesis. Library, Iowa State University of Science and Technology, Ames, Iowa.

Fox, L. and Parker, I. B. 1968. Chebyshev polynomials in numerical analysis. New York, N.Y., Oxford University Press.

Ghizzetti, A. and Ossicini, A. 1970. Quadrature formulae. Switzerland, Birkhäuser Verlag Basel Und Stuttgart.

Guest, P. G. 1958. The spacing of observations in polynomial regression. Annals of Mathematical Statistics 29: 294-299.

Hahn, G. J. and Shapiro, S. S. 1967. Statistical Models in Engineering. New York, N.Y., John Wiley & Sons, Inc.

Hildebrand, F. B. 1956. Introduction to numerical analysis. New York, N.Y. McGraw-Hill Book Company, Inc.

Hoel, P. G. 1958. Efficiency problems in polynomial estimation. Annals of Mathematical Statistics 29: 1134-1146.

Hoel, P. G. 1965a. Minimax designs in two dimensional regressions. Annals of Mathematical Statistics 36: 1097-1106.

Hoel, P. G. 1965b. Optimum designs for polynomial extrapolations. Annals of Mathematical Statistics 36: 1483-1493.

Hoel, P. G. and Levine, A. 1964. Optimal spacing and weighing in polynomial prediction. Annals of Mathematical Statistics 35: 1553-1560.

Hotelling, H. 1944. Some improvements in weighing and other experimental techniques. Annals of Mathematical Statistics 15: 297-306.

Isaacson, E. and Keller, H. B. 1966. Analysis of numerical methods. New York, N.Y., John Wiley & Sons, Inc.

Karlin, S. and Studden, W. J. 1966. Tchebycheff Systems: with applications in analysis and statistics. New York, N.Y., Interscience.

Kempthorne, O. 1947. A simple approach to confounding and fractional replication in factorial experiments. Biometrika 34: 255-272.

Kempthorne, O. 1952. The design and analysis of experiments. New York, N.Y., John Wiley & Sons, Inc.

Kiefer, J. 1959. Optimum experimental designs. Journal of the Royal Statistical Association, Series B, 21: 273-319.

Kiefer, J. and Wolfowitz, J. 1960. The equivalence of two extremum problems. Canadian Journal of Mathematics 12: 363-366.

Lentner, N. M. 1969. Generalized least squares estimation of a subvector of parameters in randomized fractional experiments. Annals of Mathematical Statistics 40: 1344-1352.

Plackett, R. L. and Burman, J. P. 1946. The design of optimum multifactorial experiments. Biometrika 33: 305-325.

Rao, C. R. 1946. On hypercubes of strength d and a system of confounding in factorial experiments. Bulletin of the Calcutta Mathematical Society 38: 67-78.

Rao, C. R. 1947. Factorial experiments derivalbe from combinatorial arrangements of arrays. Journal of the Royal Statistical Society (supplement) 9: 128-139.

Sansone, G. 1959. Orthogonal functions. New York, N.Y., Interscience.

Satterthwaithe, F. E. 1959. Random balance experimentation. Technometrics 1: 111-138.

Shah, B. V. and Kempthorne, O. 1962. Randomization in fractional factorials. Unpublished monograph. Statistical Laboratory, Iowa State University of Science and Technology, Ames, Iowa.

Timman, A. F. 1963. Theory of approximation of functions of real variable. New York, N.Y., The Macmillan Company.

Tippett, L. C. H. 1934. Applications of statistical methods to the control of quality in industrial production. Manchester Statistical Society.

Todd, J. 1963. Introduction to the constructive theory of functions. Switzerland, Birkhauser Verlag Basel und Stuttgart.

Wald, A. 1943. On the efficiency of statistical investigations. Annals of Mathematical Statistics 14: 134-140.

Wendroff, B. 1966. Theoretical numerical analysis. New York, N.Y., Academic Press.

White, R. F. 1963. Randomization analysis of the general experiment. Unpublished Ph.D. thesis. Library, Iowa State University of Science and Technology, Ames, Iowa.

Whittaker, E. Y. and Watson, G. N. 1934. Modern Analysis. England, Cambridge University Press.

Whitwell, J. C. and Morbey, G. K. 1961. Reduced designs of resolution five. Technometrics 3: 459-478.

Yates, F. 1933. Complex experiments. Journal of the Royal Statistical Society (supp.) 2: 181-247.

Zyskind, G. 1962. On structure, relations, and expectations of mean squares. Sankya, A, 24: 115-148.

## IX. ACKNOWLEDGMENTS

I wish to express my deep gratitude to my advisor Dr. Oscar Kempthorne and to Dr. George Zyskind for their patience, encouragement and guidance since I came to Iowa State University until the completion of this thesis.

I am also very grateful to my parents, especially my mother Zenaide, for their continuous encouragement and confidence since my early school years, and to my wife Marly for her help during my graduate studies.

I would like to thank the Coordenacao para Aperfeicoamento de Pessoal do Ensino Superior, (CAPES), and the Organization of American States for sponsoring my initial stages of graduate studies, Dr. Leopoldo Nachbin, Dr. Madan Lal Puri, and Dr. Edmund A. Gehan for their encouragement, and Dr. Stephen George for aid in the writing of the thesis.

Part of this work was accomplished while employed as an Instructor and Research Associate at the Department of Biomathematics, University of Texas M.D. Anderson Hospital and Tumor Institute at Houston, and partial support from Grant No. CA 12014 to the Southwest Cancer Chemotherapy Study Group is acknowledged.

Finally, it is appropriate to acknowledge support at Iowa State University for the major part of the research reported in this thesis under Contract No. F33615-68-C-1343, monitored by Aerospace Research Laboratories, Wright-Patterson

Air Force Base, Ohio, and in the terminal stages under National Science Foundation Grant GP-24614, directed by Professors Kempthorne and Zyskind.