OPTIMUM DESIGNS IN REGRESSION PROBLEMS

By J. Kiefer[1] and J. Wolfowitz[2]

Cornell University

1. Introduction and Summary. Although regression problems have been considered by workers in all sciences for many years, until recently relatively little attention has been paid to the optimum design of experiments in such problems. At what values of the independent variable should one take observations, and in what proportions? The purpose of this paper is to develop useful computational procedures for finding optimum designs in regression problems of estimation, testing hypotheses, etc. In Section 2 we shall develop the theory for the case where the desired inference concerns just one of the regression coefficients, and illustrative examples will be given in Section 3. In Section 4 the theory for the case of inference on several coefficients is developed; here there is a choice of several possible optimality criteria, as discussed in [1]. In Section 5 we treat the problem of global estimation of the regression *function*, rather than of the individual coefficients.

We shall now indicate briefly some of the computational aspects of the search for optimum designs by considering the problem of Section 2 wherein the inference concerns one of k regression coefficients. For the sake of concreteness, we shall occasionally refer here to the example of polynomial regression on the real interval $[-1, 1]$, where all observations are independent and have the same variance. The quadratic case is rather trivial to treat by our methods, so we shall sometimes refer here to the case of cubic regression. In the latter case we suppose all four regression coefficients to be unknown, and we want to estimate or test a hypothesis about the coefficient a_3 of x^3. If a fixed number N of observations is to be taken, we can think of representing the proportion of observations taken at any point x by $\xi(x)$, where ξ is a probability measure on $[-1, 1]$. To a first approximation (which is discussed in Section 2), we can ignore the fact that in what follows $N\xi$ can take only integer values. We consider three methods of attacking the problem of finding an optimum ξ:

A. The direct approach is to compute the variance of the best linear estimator of a_3 as a function of the values of the independent variable at which observations are taken or, equivalently, as a function of the moments of ξ. Denoting by μ_i the ith moment of ξ, and assuming ξ to be concentrated entirely on more than three points (so that a_3 is estimable), we find easily that the *reciprocal* of

Received April 21, 1958; revised November 25, 1958.

[1] Research under contract with the Office of Naval Research.

[2] The research of this author was supported by the U. S. Air Force under Contract No. AF 18(600)-685, monitored by the Office of Scientific Research.

this variance is proportional to

$$\frac{\mu_5^2(\mu_1^2 - \mu_2) + 2\mu_5(\mu_2^2\mu_3 + \mu_3\mu_4 - \mu_1\mu_3^2 - \mu_1\mu_2\mu_4) - \mu_4^3 + \mu_4^2(\mu_2^2 + 2\mu_1\mu_3) - 3\mu_4\mu_2\mu_3^2 + \mu_3^4}{\mu_4(\mu_2 - \mu_1^2) - \mu_3^2 - \mu_2^3 + 2\mu_1\mu_2\mu_3} + \mu_6$$

in the case of cubic regression.

The problem is to find a ξ on $[-1, 1]$ which maximizes this expression. Thus, this direct approach leads to a calculation which appears quite formidable. This is true even if one uses the remark on symmetry of the next paragraph and restricts attention to symmetrical ξ, so that $\mu_i = 0$ for i odd. For polynomials of higher degree or for regression functions which are not polynomials, the difficulties are greater.

B. The results of Section 2 yield the following approach to the problem: Let $c_0 + c_1x + c_2x^2$ be a best Chebyshev approximation to x^3 on $[-1, 1]$, i.e., such that the maximum over $[-1, 1]$ of $|x^3 - (c_0 + c_1x + c_2x^2)|$ is a minimum over all choices of the c_i, and suppose B is the subset of $[-1, 1]$ where the maximum of this absolute value is taken on. Then ξ must give measure one to B, and the weights assigned by ξ to the various points of B (there are four in this case) can be found either by solving the *linear* equations (2.10) or by computing these weights so as to make ξ a maximin strategy for the game discussed in Section 2. Two points should be mentioned:

(1) In the general polynomial case, where there are k parameters ($k = 4$ here), the results described in [10], p. 42, or in Section 2 below imply that there is an optimum ξ concentrated on at most k points. Thus, even if we use this result with the approach of the previous paragraph, we obtain the following comparison in a k-parameter problem in Section 2:

Method A: minimize a nonlinear function of $2k - 1$ real variables.

Method B: solve the Chebyshev problem and then solve $k - 1$ simultaneous *linear* equations.

The fact that the solution of the Chebyshev problem can often be found in the literature (e.g., [2]) makes the comparison of the second method with the first all the more favorable.

(2) Although the computational difficulty cannot in general be reduced further, in the case of polynomial regression on $[-1, 1]$ there is present a kind of symmetry (discussed in Section 2) which implies that there is an optimum ξ which is symmetrical about 0 and which is concentrated on four points; thus, in the case of cubic regression, this fact reduces the computation under Method A to a minimization in 3 variables, but Method B involves only the solution of a single linear equation.

C. A third method, which rests on the game-theoretic results of Section 2, and which is especially useful when one has a reasonable guess of what an optimum ξ is, involves the following steps: first guess a ξ, say ξ^*, and compute the minimum on the left side of (2.8); second, if this minimum is achieved for $c = c^*$, compute the square of the maximum on the right side of (2.9); then, if these two computations yield the same number, ξ^* is optimum. If one has a guess of a class of ξ's depending on one or several parameters, among which it is thought that there is an optimum ξ, then one can maximize over that class at the end of the first step and, the maximum being at ξ^*, go through the same analysis as above. This method is illustrated in Example 3.5 and Example 4. Of course, the remarks (1) and (2) of the previous paragraph can be used in applying Method C, as in these examples.

In the example of cubic regression just cited, the optimum procedure turns out to be $\xi(-1) = \xi(1) = \frac{1}{6}$, $\xi(\frac{1}{2}) = \xi(-\frac{1}{2}) = \frac{1}{3}$. It is striking that any of the commonly used procedures which take equal numbers of observations at equally spaced points on $[-1, 1]$ requires over 38% more observations than this optimum procedure in order to yield the same variance for the best linear estimator of a_3 (see Example 3.1); the comparison is even more striking for higher degree regression. The unique optimum procedure in the case of degree h is given by (3.3).

The comparison of a direct computational attack, analogous to that of A above, with the methods developed in Sections 4 and 5 for the problems considered there, indicates even more the inferiority of the direct attack. In particular cases, e.g., Example 5.1, special methods may prove useful.

Among recent work in the design of experiments we may mention the papers of Elfving [3], [4], Chernoff [5], Williams [11], Ehrenfeld [12], Guest [13], and Hoel [15]. Only Guest and Hoel explicitly consider computational problems of the kind discussed below. Our methods of employing Chebyshev and game theoretic results seem to be completely new. The results obtained in the examples below are also new, except for some slight overlap with results of [13] and [15], which is explicitly described below.

We shall consider elsewhere some further problems of the type considered in this paper.

2. The optimum design relative to 1 out of k regression coefficients. Let $f_1, \cdots, f_k$ be k real-valued functions on a given space $\mathfrak{X}$. Throughout this section we assume a topology is given on $\mathfrak{X}$ in which

(2.1) $$\mathfrak{X} \text{ is compact}; f_1, \cdots, f_k \text{ are continuous.}$$

We also assume

(2.2) $$f_1, \cdots, f_k \text{ are linearly independent on } \mathfrak{X}.$$

Since we will be considering a regression problem in which the f_i are known functions and $\sum_i a_i f_i$ is the regression function, (2.2) is really only an assumption of identifiability of the a_i which will avoid trivial circumlocutions. Without some assumption like the first part of (2.1), there may trivially exist procedures which estimate some of the regression coefficients with arbitrarily small variance, as can be seen in the example of estimation of the slope of a straight line on $\mathfrak{X}$ = real line. The assumption of continuity of the f_i can be somewhat weakened, as will be clear from our proofs.

We consider the following regression setup: For any point x (value of the independent variable) in $\mathfrak{X}$, one can observe a random variable Y_x for which

$$EY_x = \sum_1^k a_i f_i(x), \tag{2.3}$$
$$\operatorname{Var}(Y_x) = \sigma^2,$$

where $a = (a_1, \cdots, a_k)$ is the vector of regression coefficients, an unknown element of $\mathfrak{A}$. The value of σ^2 will usually be unknown. (The case where σ^2 can depend on x in a way which is known except for a proportionality constant will be discussed in the last paragraph of this section.) An integer n is given (usually $n > k$), and the experimenter must select a collection $\mathbf{X} = (x_1, \cdots, x_n)$ of n points in $\mathfrak{X}$ at which the independent random variables $Y_{x_1}, \cdots, Y_{x_n}$ are to be observed. The x_i need not be distinct, but if $i \neq j$ and $x_i = x_j$ we shall still, without confusion, write Y_{x_i} and Y_{x_j} for two *independent* random variables.

Any X can be viewed as a measure η on $\mathfrak{X}$ which assigns to each point x a mass equal to the number of x_i in X which are equal to x. Dividing this measure by n, we obtain a discrete probability measure ξ on $\mathfrak{X}$ which assigns to each point of $\mathfrak{X}$ a measure equal to an integral multiple of $1/n$. In the present section (a similar discussion applying in Sections 4 and 5), we shall be concerned with choosing a ξ (hence, an X) to maximize a quantity of the form

$$\min_c \int_{\mathfrak{X}} H_c(x)\eta(dx) = \min_c \int_{\mathfrak{X}} nH_c(x)\xi(dx), \tag{2.4}$$

where the form of H_c is determined by the problem at hand. The fact that ξ can only take on multiples of $1/n$ as its values makes this problem of maximization quite unwieldy in general. We shall treat, instead, a problem whose solution will sometimes give a solution to the original problem and which will usually give a good approximation to the latter: *Find a probability measure ξ^* on $\mathfrak{X}$ for which the right side of* (2.4) *is a maximum*: i.e., we maximize (2.4) with no restriction on ξ. Of course, the maximum does not depend on n. Thus, if n is such that $n\xi^*$ takes on only integral values, this yields an exact solution η to the original problem. We shall see in Sections 3 and 5 that, in two typical examples, ξ^* takes on only values which are multiples of $1/(2k - 2)$ (Example 3.1) or $1/k$ (Example 5.1), so that this situation is not vacuous. Moreover, there will typically be a ξ^* which is concentrated on approximately k points; thus, when $n\xi^*$ does not take on only integral values, obvious integral approximations η' to $n\xi^*$ will yield values of (2.4) whose ratio to the maximum tends to 1 as $n \to \infty$ (it is easy to give a bound on the difference of this ratio from unity). Thus, the characterization of a single ξ^* which yields an *almost* optimum design for all large n, in distinction to finding the best ξ which may depend in a complicated fashion on n, seems to be of practical value.

We therefore define Ξ to be the space of all discrete probability measures ξ on $\mathfrak{X}$. We could, more generally, specify a Borel field $\mathfrak{B}$ on $\mathfrak{X}$ and let Ξ be the class of all measures ($\mathcal{B}$) on $\mathfrak{X}$; however, in all of our applications (see Theorem 2) it will suffice to let $\mathcal{B}$ consist of countable sets and their complements.

In the present section we are concerned with statistical inference about the single parameter a_k, where all a_i are assumed unknown. We shall give a precise definition of optimality in the next paragraph. What this definition means is that we restrict ourselves to designs for which a_k is estimable (i.e., for which there exist linear unbiased estimators of a_k; in practice, of course, n will have to be suitably large for there to exist such designs), and seek a design for which the linear unbiased estimator of a_k with minimum variance (best linear estimator, or b.l.e.) has a variance which is a minimum over all designs, within the approximation noted two paragraphs above. It is well known that such a design is optimum for problems of point estimation of a_k if the Y_x are assumed to be normal, in the sense that (for example) it yields a minimax procedure for any of a wide variety of weight functions; when the distributions of Y_x are assumed to belong to any larger class, the same result holds for the squared error loss function. For problems of interval estimation and hypothesis testing or m decisions, similar optimality results hold under normality if σ^2 is known. If σ^2 is unknown, such results hold provided every design for which a_k is estimable yields as many degrees of freedom to error as does the design we obtain; see Example 3.4 in Section 3 for further discussion.

We now define precisely the term "optimum" as used in this section. There are a few preliminaries. In the original description of a design, let X be a design for which a_k is estimable. Let $h_1, \cdots, h_{k-1}$ be numbers such that the function $f_k^* = f_k - \sum_1^{k-1} h_j f_j$ is orthogonal to f_i for $i < k$ in the sense that

$$\sum_{r=1}^{n} f_i(x_r) f_k^*(x_r) = 0, \qquad i < k. \tag{2.5}$$

Let $a^* = (a_1^*, \cdots, a_k^*)$ be such that $\sum_1^k a_i f_i = \sum_1^{k-1} a_i^* f_i + a_k^* f_k^*$; thus, $a_k^* = a_k$. For the least squares setup in terms of a^*, the orthogonality of f_k^* to the f_i for $i < k$ makes the last of the normal equations

$$\sum_{r=1}^{n} [f_k^*(x_r)]^2 a_k^* = \sum_{r=1}^{n} f_k^*(x_r) Y_{x_r}, \tag{2.6}$$

so that σ^2 times the *reciprocal* of the variance of the b.l.e. of $a_k^* = a_k$ is $\sum_r [f_k^*(x_r)]^2$. Since f_k^* is orthogonal to $f_1, \cdots, f_{k-1}$, this last sum is just the square of the distance of the n-vector $(f_k(x_1), \cdots, f_k(x_n))$ from the linear space spanned by the vectors $(f_i(x_1), \cdots, f_i(x_n))$ for $i < k$, namely,

$$\min_c \sum_r [f_k(x_r) - \sum_{j=1}^{k-1} c_j f_j(x_r)]^2, \tag{2.7}$$

where we have written c for $(c_1, \cdots, c_{k-1})$. Since (2.7) is σ^2 times the inverse of the variance of the b.l.e. of a_k, a design X will minimize that variance if it maximizes (2.7). Thus, finally, in terms of the probability measures ξ we have introduced above, we make the following

DEFINITION. *A measure ξ^* in Ξ is said to be an optimum design (for the pa-*

rameter a_k) if

$$\min_c \int [f_k(x) - \textstyle\sum_1^{k-1} c_j f_j(x)]^2 \xi^*(dx) \tag{2.8}$$
$$= \max_{\xi \in \Xi} \min_c \int [f_k(x) - \textstyle\sum_1^{k-1} c_j f_j(x)]^2 \xi(dx).$$

For any ξ in Ξ, the ratio of the left side of (2.8) (*with ξ for ξ^**) *to the right will be called the efficiency $e(\xi)$ of ξ.*

Of course, the practical meaning of efficiency is that, if one design has r times the efficiency of the second design, then the latter requires r times as many observations as the former in order to obtain the same value for the left side of (2.4). We note that it is a consequence of this definition that an optimum design is optimum for all values of σ^2.

The form of (2.8) is very suggestive of a game, and we shall exploit that fact presently. However, the main aspect of our technique for computing an optimum ξ^* has nothing to do with the game formulation, so we treat that aspect first. Our technique is to throw the main computational difficulties into a Chebyshev approximation problem, which can often be solved by standard methods and which, for many important $\{f_j\}$, even has a solution which can be found in the literature. We shall call $c^* = (c_1^*, \cdots, c_{k-1}^*)$ a *Chebyshev coefficient vector* if $\sum_1^{k-1} c_j^* f_j$ is a best approximation to f_k on $\mathfrak{X}$ in the sense of Chebyshev, i.e., in the uniform norm:

$$\min_c \max_{x \in \mathfrak{X}} |f_k(x) - \sum_1^{k-1} c_j f_j(x)| = \max_{x \in \mathfrak{X}} |f_k(x) - \sum_1^{k-1} c_j^* f_j(x)|. \tag{2.9}$$

Let $m(c^*)$ denote the right side of (2.9), and let $B(c^*)$ be the set of points x for which $| f_k(x) - \sum_1^{k-1} c_j^* f_j(x)| = m(c^*)$. Our first result gives a simple geometric sufficient condition for a ξ to be optimum; this is valid even without the conditions that yield the game-theoretic results of Theorem 2.

THEOREM 1. *If c^* is Chebyshev and $\xi(B(c^*)) = 1$ and*

$$\int [f_k(x) - \textstyle\sum_1^{k-1} c_j^* f_j(x)] f_i(x) \xi(dx) = 0 \tag{2.10}$$

for $i < k$, then ξ is optimum.

PROOF: According to (2.10), $\sum_1^{k-1} c_j^* f_j$ is the projection relative to ξ of f_k on the linear space spanned by $f_1, \cdots, f_{k-1}$. Hence, for any element ξ' of Ξ,

$$\begin{aligned} &\min_c \int [f_k(x) - \textstyle\sum_1^{k-1} c_j f_j(x)]^2 \xi(dx) \\ &= \int [f_k(x) - \textstyle\sum_1^{k-1} c_j^* f_j(x)]^2 \xi(dx) \\ &= [m(c^*)]^2 \geqq \int [f_k(x) - \textstyle\sum_1^{k-1} c_j^* f_j(x)]^2 \xi'(dx) \\ &\geqq \min_c \int [f_k(x) - \textstyle\sum_1^{k-1} c_j f_j(x)]^2 \xi'(dx), \end{aligned} \tag{2.11}$$

which proves the desired result.

The question arises as to whether there always exists a ξ which satisfies the hypotheses of Theorem 1 and whether, in fact, the conditions of the theorem are also *necessary* for a ξ to be optimum. There also arises the question of whether we can find a useful bound such that there is an optimum ξ which assigns positive probability to at most the number of points given by this bound. These questions can be answered directly algebraically, but since the results we require already appear in the literature in connection with the analysis of certain games, we shall therefore consider the following *zero-sum two-person game associated with the design problem*: player 1 (resp., 2) has $\mathfrak{X}$ (resp., C = Euclidean $(k - 1)$-space) as his space of pure strategies; the payoff function is $K(x, c) = [f_k(x) - \sum_1^{k-1} c_j f_j(x)]^2$; the space of mixed strategies of player 1 is Ξ, while that of player 2 is immaterial, since the convexity of K in C implies, according to Jensen's inequality, that for any randomized strategy of player 2 there is a nonrandomized strategy which is at least as good for all x. Of course the important thing is that an optimum (maximin) strategy for player 1 represents an optimum design. We now state the simple modifications of certain results of [6] which we require.

LEMMA. *The game of Ξ vs. C is determined, player 2 has a nonrandomized minimax strategy c^*, and player 1 has a maximin strategy ξ^* which is concentrated on at most $k - p$ points, where p is the dimensionality of the convex set of nonrandomized minimax strategies of player 2.*

PROOF: Let C_N be the set of all c for which $c'c = \sum_1^{k-1} c_i^2 \leqq N^2$, and let $\bar{C}_N$ be the complement of C_N. Since the f_i are linearly independent, there is a finite subset H of $\mathfrak{X}$ such that, for every c with $c'c = 1$, $\sum_1^{k-1} c_i f_i(x)$ is nonzero for at least one x in H. Hence, if ξ' assigns positive probability to each x in H, we clearly have $|\sum_1^{k-1} c_i \sum_{x \varepsilon H} f_i(x)\xi'(x)| > \epsilon > 0$ for all c such that $c'c = 1$, and thus this absolute value is $> N\epsilon$ for $c'c = N^2$. Since f_k is bounded, we conclude that $\inf_{c \varepsilon \bar{C}_N} K(\xi', c) \to \infty$ as $N \to \infty$. Hence, there is an N' such that for any c in $\bar{C}_{N'}$ there is a c' in $C_{N'}$ with $\sup_\xi K(\xi, c') < \sup_\xi K(\xi, c)$. Thus c^* is minimax if and only if $c^* \varepsilon C_{N'}$ and c^* is minimax when the space of player 2 is restricted to $C_{N'}$. Since C_N is compact and K is continuous, the game of Ξ vs. C_N is determined, and there exists, for all $N > N'$, a minimax strategy c^* which we can take to be a fixed member of $C_{N'}$. Let p be the dimension of the (convex) set of such minimax strategies in $C_{N'}$. There also exists a maximin strategy ξ_N^* for the game of Ξ vs. C_N, and by [6] we can for $N > N'$ take ξ_N^* to be concentrated on at most $k - p$ points. Let $\xi_j = [(j - 1)\xi_j^* + \xi']/j$. Clearly, for each j there is an N_j such that $K(\xi_j, c^*) < K(\xi_j, c)$ for all c in $\bar{C}_{N_j}$. Thus, since ξ_j^* is maximal with respect to c^*, we have, for $N_j' > N'$,

$$
\begin{aligned}
\sup_{\xi \varepsilon \Xi} \inf_{c \varepsilon C} K(\xi, c) &\geqq \inf_{c \varepsilon C} K(\xi_j, c) = \inf_{c \varepsilon C_{N_j}} K(\xi_j, c) \\
&\geqq \left(1 - \frac{1}{j}\right) \inf_{c \varepsilon C_{N_j}} K(\xi_j^*, c) = \left(1 - \frac{1}{j}\right) K(\xi_j^*, c^*) \\
&= \left(1 - \frac{1}{j}\right) \sup_\xi K(\xi, c^*) \geqq \left(1 - \frac{1}{j}\right) \inf_{c \varepsilon C} \sup_\xi K(\xi, c).
\end{aligned}
\tag{2.12}
$$

Letting $j \to \infty$, we see that the game of Ξ vs. C is determined, that c^* is minimax, and that if $\{\xi^*_{j_i}\}$ is a subsequence of the $\{\xi^*_j\}$ which converges to a limit ξ^* which is concentrated on no more than $k - p$ points (such a subsequence and limit exist, by the compactness of $\mathfrak{X}$) and c'' minimizes $K(\xi^*, c)$, we have

$$\begin{aligned}\sup_{\xi} \inf_{c \varepsilon C} K(\xi, c) &= \lim_{i\to\infty} \inf_{c \varepsilon C} K(\xi_{j_i}, c) \leqq \lim_{i\to\infty} K(\xi_{j_i}, c'') \\ &= K(\xi^*, c'') = \inf_{c \varepsilon C} K(\xi^*, c),\end{aligned} \tag{2.13}$$

so that ξ^* is maximin. Thus, the lemma is proved.

We mention in passing several other related points: The bound $k - p$ is indicated in [6] not to be the best possible and is reduced under conditions (c^* in the boundary of a compact C) for which it is difficult to find general counterparts here. Also, it is evident that c^* is unique ($p = 0$) if $K(x, c)$ is strictly convex in c, but strict convexity is clearly not a useful condition in our problem. If $\mathfrak{X}$ is not compact or the f_j are not continuous, suitable assumptions will still imply determinateness, but the other results will have to be stated in terms of ϵ-optimum strategies.

The above lemma indicates one method for trying to compute a ξ^*: For simplicity, assume $p = 0$ or that we have no knowledge of p. The ξ's on $\mathfrak{X}$ which are concentrated on at most k points form a $(2k - 1)$-parameter family. One can thus, in principle, maximize $\min_c K(\xi, c)$ with respect to these $2k - 1$ parameters and obtain an optimum ξ^*. As we have indicated in the introduction, this is usually an unrewarding task, and the method indicated in Theorem 1 seems far superior in practical examples. The consequences of the lemma for the method of Theorem 1 may be summarized as follows:

THEOREM 2. *If ξ is maximal with respect to c^* while c^* is minimal with respect to ξ, then ξ is optimum and c^* is Chebyshev. Every optimum ξ satisfies the conditions of Theorem 1 for every Chebyshev c^*. There exists an optimum ξ concentrated on at most $k - p$ points, where p is the dimensionality of the Chebyshev vectors.*

PROOF: The Chebyshev vectors clearly coincide with the minimax strategies. If ξ is maximin and c^* is minimax, then determinateness implies that c^* is minimal with respect to ξ, i.e., $\min_c K(\xi, c) = K(\xi, c^*)$. Thus, $\sum_1^{k-1} c_j^* f_j$ is the projection, relative to ξ, of f_k on the linear space spanned by $f_1, \cdots, f_{k-1}$, so that (2.10) clearly holds. Since, by (2.11), max min $K(\xi, c) = [m(c^*)]^2$ is the value of the game, $\xi(B_{c_*}) = 1$. The last assertion of the theorem is taken directly from the lemma, while the first is a general result in the theory of games. We note that any optimum ξ must give measure one to the intersection of all $B(c^*)$ for c^* Chebyshev.

We have mentioned, in Theorem 1 and in the second paragraph below the proof of the lemma, two computational approaches. The first sentence of Theorem 2 indicates a useful approach if one can make a good guess of ξ: guess a ξ' and compute $\min_c K(\xi', c) = K(\xi', c')$ (say); compute $\max_x K(x, c')$; if these two are equal, then ξ' is optimum. This is an approach which is standard in game theory and which has proved useful in many examples; it sometimes helps to let ξ' depend on a few parameters, with respect to which one maximizes

$\min_c K(\xi', c)$. A comparison of the various methods for obtaining an optimum ξ was given in an example in Section 1.

In the next section we shall give several examples of the computation of optimum ξ's. We shall not bother to list in detail all of the standard results in approximation theory which are useful in such computations. We mention here for future reference only the classical generalized Chebyshev theorem [2, p. 74], which states that if $\mathfrak{X}$ is a compact real interval and if no nontrivial linear combination of $f_1, \cdots, f_{k-1}$ has more than $k - 2$ zeros (in this case, these f_i are called a *Chebyshev system*), then the Chebyshev vector c^* is unique and is characterized by the fact that there are at least k points at which $f_k - \sum_1^{k-1} c_j^* f_j$ attains its maximum absolute deviation from zero, the maximum being taken on with successive alternations in sign. (The literature contains generalizations of this result to other spaces.)

Before proceeding further it is relevant here to point out the following connections with earlier results:

1) Elfving [3] considered the special case where $\mathfrak{X}$ contains a finite number of discrete points. It follows from his elegant geometrical argument that the optimum ξ is concentrated on at most k points and satisfies (2.10).

2) Consider the case of polynomial regression ($\mathfrak{X}$ a closed interval of the real line, $f_i(x) = x^{i-1}$). Then $p = 0$ by the Chebyshev theorem cited above. Theorem 2 then says, inter alia, that there exists an optimum ξ concentrated on at most k points. This result (for this important particular case) is already well known in the theory of moment problems ([10], p. 42). It holds *identically in* σ^2. If it did not hold for all σ^2 it would be useless in our problem when σ^2 is unknown. This result holds even when (2.18) below is true, with fixed v.)

We now give a simple result on the uniqueness of the optimum ξ^*.

THEOREM 3. *If $\mathfrak{X}$ is a compact real interval, $f_1, \cdots, f_{k-1}$ is a Chebyshev system, and $B(c^*)$ contains exactly k points, then the optimum ξ^* is unique.*

PROOF: Let $x_1, \cdots, x_k$ be the ordered members of $B(c^*)$, and let Q be the $(k - 1) \times k$ matrix whose (i, j)th element is $(-1)^j f_i(x_j)$. Let ξ denote a k-vector whose jth component is the number $\xi(x_j)$. According to (2.10), which, by Theorem 2, is necessary, and the Chebyshev theorem cited above, any optimum ξ must satisfy

$$Q\xi = 0. \tag{2.14}$$

(Of course, it must also satisfy $\xi(B(c^*)) = 1$.) Now, Q has rank $k - 1$, since, if it had smaller rank, a nontrivial weighted sum of rows of Q would be 0 and the f_i could not be a Chebyshev system. The linear equations (2.14) thus have a one-dimensional set of solutions ξ, and clearly at most one of these can be a probability measure. This completes the proof.

If $B(c^*)$ consists of more than k points, an analysis like that above will give information on how large the class of optimum ξ's can be.

Remark on symmetry (invariance): As we have indicated in Section 1, it will sometimes be easy, as in the case of polynomial regression, to infer that there is an optimum ξ with some symmetry property. Formally, suppose that there is a group G of transformations on $\mathfrak{X}$ such that for each g in G there is a transformation g' on $\mathfrak{A}$ such that, writing $(g'a)_i$ for the ith coordinate of $g'a$, we have $(g'a)_k = a_k$ for g in G and

$$(2.15) \qquad \sum_i a_i f_i(x) = \sum_i (g'a)_i f_i(gx)$$

for all x and all $(a_1, \cdots, a_k)$. (One may let g' act on the vector of functions f_i instead of on $\underset{\sim}{a}$.) Then the problem in terms of the parameters $(g'a)_i$ and the independent variable gx coincides with the original problem. Hence, if ξ is optimum for the original problem, it is also optimum for the above problem in terms of gx and hence the measure ξ_g defined by

$$(2.16) \qquad \xi_g(A) = \xi(g^{-1}A)$$

is optimum for the original problem in terms of x. Suppose for the moment that G contains a finite number, say L, of elements. Write

$$(2.17) \qquad \bar{\xi} = \sum_{g \varepsilon G} \xi_g / L.$$

It is easy to prove that, if ξ is optimum, then so is $\bar{\xi}$; in fact, this is obvious statistically, since the variance of the average of L b.l.e.'s from the L independent experiments ξ_g with N/L observations each cannot be less than that of the b.l.e. from $\bar{\xi}$ with N observations (since $\bar{\xi}$ can be broken up into such experiments), but is clearly equal to the variance of the b.l.e. from ξ based on N observations. Thus, we have:

There exists an optimum design which is symmetric with respect to (invariant under) G.

The analogous result can be proved for G compact or satisfying conditions which yield the usual minimax invariance theorem in statistics; see, e.g., [7].

The fact that there exists an optimum symmetric design and an optimum design concentrated on (e.g.) k points does not imply the existence of an optimum design with both of these properties. For example, if $\mathfrak{X} = [-1, 1]$, $k = 2$, $f_1(x) = 1$, and $f_2(x) = x^2$, there is an optimum design concentrated on the two points 0 and 1, but the only symmetric design requires the three points 0, -1, and 1. However, in the event that g' does not act (as it does in the example just cited) as the identity for every g, we may be able to obtain some simplification. For example, without discussing the most general possibility, let us suppose that Q is a set of integers containing k and such that $(g'a)_j = a_j$ for all g if $j \varepsilon Q$, while $\sum_g (g'a)_j = 0$ for j not in Q. Consider the problem of finding an optimum design ξ on the space of equivalence classes of $\mathfrak{X}$ under the equivalence $x \sim x'$ if $x' = gx$ for some g, where the regression function is $\sum_{j \varepsilon Q} a_j f_j(x)$ (at the equivalence class of x). If there are q integers in Q, there is by Theorem 2 an optimum τ^* concentrated on at most q points. This τ^* corresponds to a unique symmetric (with respect to G) measure ξ^* on $\mathfrak{X}$, and it is easy to see that (2.10) is satisfied for *all* $i < k$. Thus, if there are L elements in G, this ξ^* is concentrated on at most qL points. For example, in the case of polynomial regression of even degree h $(= k - 1)$ on $[-1, 1]$, G contains two elements and the set Q corresponds to the $q = 1 + h/2$ even powers, and we obtain that there is a symmetric optimum ξ concentrated on at most $h + 2$ points. The actual case (see Ex. 3.1) is that there is a symmetric optimum ξ concentrated on $k = h + 1$ points; the previous argument did not give the best result because τ^* gave positive probability to the equivalence class of 0, which corresponds to only one point of $\mathfrak{X}$. The best result could, however, be obtained using another argument: since, according to Theorem 3, the optimum ξ is *unique*, our discussion of two paragraphs above implies that it must be symmetric, and it is thus concentrated on $h + 1$ points. Similarly, one could conclude that there is a symmetric optimum design concentrated on $h + 1$ points when h is odd, either by using Theorem 3, or else by invoking an obvious modification of the previous argument for the case when $(g'a)_k = \pm a_k$. A similar result holds in the setup of Ex. 3.5.

Remark on heteroscedasticity and variable cost: Suppose the second line of (2.3) is replaced by

$$\text{Var}(Y_x) = [v(x)\sigma]^2, \tag{2.18}$$

where v is a known positive continuous function on $\mathfrak{X}$. To avoid trivialities, assume $v(x)$ bounded away from 0. Then, replacing Y_x by $Y_x^* = Y_x/v(x)$ and $f_i(x)$ by $f_i^*(x) = f_i(x)/v(x)$, it is clear that the entire discussion of this section goes through exactly as before (i.e., assuming (2.3)) since the a_i for which $EY_x = \sum a_i f_i(x)$ are the same a_i as those for which $EY_x^* = \sum a_i f_i^*(x)$, and the latter setup satisfies the original condition (2.3) of this section.

If there is a cost $c(x)$ of taking an observation at the point x, and the total cost rather than the total number of observations is to be kept constant, it is easily seen that an optimum design is obtained by going through the analysis of this section with $v(x)$ above replaced by $v(x)[c(x)]^{1/2}$.

Similar remarks will apply to the problems considered in Sections 4 and 5.

3. Examples of optimum designs in the case of Section 2.

Example 3.1. *Polynomials on* $[\alpha, \beta]$. One of the most important practical examples is that where $\mathfrak{X}$ is the closed finite nondegenerate interval $[\alpha, \beta]$ of reals, $k = h + 1$ for some $h > 0$, and $f_j(x) = x^{j-1}$ for $1 \leqq j \leqq h + 1$; we hereafter write $b_{j-1} = a_j$, $b = (b_0, \cdots, b_h)$, $d_{j-1} = c_j$, and $d = (d_0, \cdots, d_{h-1})$. Thus, assuming that the regression function is a polynomial of degree $\leqq h$, we may want to test the hypothesis that it is actually of degree $\leqq h - 1$, i.e., that $b_h = 0$. (In Section 4 we consider the possibility of testing that the degree is $\leqq h - m$ where m is specified). We first note that we can write

$$\sum_{j=0}^{h} b_j x^j = \sum_{j=0}^{h} b_j'[(2x - \alpha - \beta)/(\beta - \alpha)]^j,$$

where $b_h' = [(\beta - \alpha)/2]^h b_h$; since $(2x - \alpha - \beta)/(\beta - \alpha)$ takes on values in $[-1, 1]$, an optimum strategy for arbitrary $[\alpha, \beta]$ is immediately obtained by an obvious change in location and scale from an optimum strategy in the case $[-1, 1]$, and we may hereafter limit our attention to the latter. Next, we note that b_h is obviously not estimable unless ξ gives positive probability to at least $h + 1$ points (of course, in practice we need $n > h + 1$ if σ^2 is unknown and $n \geqq h + 1$ if σ^2 is known). Hence, by Theorem 2 (or by the result of [10] cited in Section 2) there exists an optimum ξ concentrated on exactly $(h + 1)$ points. We shall actually find a unique ξ^* which satisfies (2.8) and gives positive probability to exactly $h + 1$ points.[3] Thus, the phenomenon concerning degrees of freedom in the estimate of σ^2 which was discussed in the sixth paragraph of Section 2, and which is illustrated in Example 3.4 below, cannot occur in the present example.

The unique Chebyshev d^* (i.e., c^*) is well known in this example: $x^h - \sum_0^{h-1} d_j^* x^j$ is simply the hth Chebyshev polynomial (see, e.g., [2]),

$$
\begin{aligned}
x^h - \textstyle\sum_0^{h-1} d_j^* x^j &= 2^{1-h}\cos(h\cos^{-1}x) \\
&= 2^{-h}\{[x + (x^2 - 1)^{1/2}]^h + [x - (x^2 - 1)^{1/2}]^h\}.
\end{aligned} \tag{3.1}
$$

Moreover, $m(d^*) = 2^{1-h}$, and this extreme value is attained in magnitude (with successive alterations in sign) by $x^h - \sum_0^{h-1} d_j^* x_j$ at the $h + 1$ points

$$
x_j = -\cos\frac{j\pi}{h}, \qquad 0 \leqq j \leqq h. \tag{3.2}
$$

Thus, $B(d^*)$ consists of these $h + 1$ points. Moreover, the above d^* is the unique Chebyshev vector, since $x^0, x^1, \cdots, x^{h-1}$ form a Chebyshev system.

According to Theorem 3, the optimum ξ^* is unique. We now show that the unique optimum ξ^* is

$$
\begin{aligned}
\xi^*(-1) &= \xi^*(1) = \tfrac{1}{2}h, \\
\xi^*\left(\cos\frac{j\pi}{h}\right) &= 1/h, \qquad 1 \leqq j \leqq h - 1.
\end{aligned} \tag{3.3}
$$

To prove this, we shall verify (2.14) for $\xi = \xi^*$, since this is just (2.10), which by Theorems 1 and 2 is necessary and sufficient for an optimum ξ. Since the d_j^*'s of (3.1) are zero if $j + h$ is odd, the polynomial of (3.1) is clearly orthogonal (with respect to ξ^*) to x^t when $t + h$ is odd. When $t + h$ is even, we can combine the weights $\xi^*(-1)$ and $\xi^*(1)$ and rewrite (2.14) as

$$
\sum_{j=0}^{h-1} (-1)^j \left(\cos\frac{\pi j}{h}\right)^t = 0. \tag{3.4}
$$

Since $\cos^t\theta$ can be written as a linear combination of $\cos t\theta$, $\cos(t - 2)\theta, \cdots$, it suffices to prove (3.4) with $\cos^t(\pi j/h)$ replaced by $\cos(rj\pi/h)$, where $h + r$

[3] For $h = 1$ and 2, the solution is given in [14]. The general solution (3.3) of the problem of Example 3.1 for a design optimum in the sense of Section 2, is also given in the abstract [11] of the apparently contemporaneous work of E. J. Williams. The methods of this author are probably different from ours because he does not seem to use probability measures ξ. The authors are indebted to H. L. Lucas for calling their attention to [11] which appeared after submission of the present manuscript.

is even and $0 \leqq r \leqq h$. But for such r we have

$$\begin{aligned}\sum_{j=0}^{h-1} (-1)^j \cos (rj\pi/h) &= \mathrm{Re}\left\{\sum_{j=0}^{h-1} \exp [ji\pi(1 + r/h)]\right\} \\ &= \mathrm{Re}\left\{\frac{1 - \exp [i\pi(h + r)]}{1 - \exp [i\pi(1 + r/h)]}\right\} = 0.\end{aligned} \tag{3.5}$$

It is interesting to compare the design ξ^* of (3.3) with the often used design $\xi^{h,M}$ (say) which assigns measure $1/M$ to each of the values $(2i - M - 1)/(M - 1)$, $i = 1, 2, \cdots, M$; thus $\xi^{h,M}$ takes an equal number of observations at each of M equally spaced points ranging from -1 to 1. Of course, $M > h$. For such a design with M observations on the interval $[0, M - 1]$, Fisher [8, p. 153] has calculated the left side of (2.4) to be $(h!)^4M(M^2 - 1)(M^2 - 4) \cdots (M^2 - h^2)/(2h)!(2h + 1)!$ To obtain the corresponding quantity for the interval $[-1, 1]$, we must divide by $[(M - 1)/2]^{2h}$, and we must divide also by M in order to obtain the left side of (2.4) with η replaced by $\xi^{h,M}$. Since $[m(d^*)]^2 = 2^{2-2h}$, we obtain for the efficiency (see the definition following (2.8)) of $\xi^{h,M}$

$$e(\xi^{h,M}) = \frac{2^{4h-2}(h!)^4}{(2h)!(2h + 1)!} \prod_{i=1}^{h} \frac{M^2 - i^2}{(M - 1)^2}. \tag{3.6}$$

The best choice of M varies: it is $h + 1$ if $h = 1$ or 2, $h + 2$ if $h = 3$, etc. For the often used procedure $\xi^{h,h+1}$, we have

$$e(\xi^{h,h+1}) = \frac{2^{4h-2}(h!)^4}{(2h)!h^{2h}(h + 1)}. \tag{3.7}$$

Of course, (3.7) becomes 1 for $h = 1$, since $\xi^{1,2} = \xi^*$ for $h = 1$; for $h = 2$, (3.7) becomes 8/9, for $h = 3$ it is 256/405 (the best procedure, $\xi^{3,5}$, has efficiency .72), etc.; for large h, by Stirling's approximation, it is approximately $\pi^{3/2}h^{1/2}2^{2h-1}e^{-2h}$, which goes to zero very rapidly. For $\xi^{h,M}$ with $M \to \infty$, the efficiency (3.6) approaches $2^{4h-2}(h!)^4/(2h)!(2h + 1)!$, which as $h \to \infty$ is approximately $\pi/8$.

To the experimenter who protests at the above comparison that the design $\xi^{h,M}$ for some $M > h$ is more to his liking than is the ξ^* of (3.3) because the former will permit him to estimate regression coefficients a_j up to a_{M-1} (instead of up to a_h), we can only answer that his problem is not the one of the present example, that he is probably using a method of inference (to "choose the polynomial of correct degree") whose properties are questionable, and that a precise statement of his decision problem would probably lead to a procedure far superior to $\xi^{h,M}$. In Sections 4 and 5 we shall consider some other related problems which may be what the experimenter is faced with, rather than the problem of the present example. The problem of "fitting the polynomial of best degree" is more unwieldy, depending strongly on the somewhat arbitrary choice of losses which are to be assigned to errors in estimation as compared with the penalty for using a polynomial of large degree.

Example 3.2. *An example where* $p > 0$. It is easy to construct examples where the p of Theorem 2 is not 0 as it is in the case of a Chebyshev system. We illustrate the situation with a very simple example. Suppose $\mathfrak{X} = [-1, 1]$, $k = 3$, $f_1(x) = 1, f_2(x) = x^2, f_3(x) = x + 1$. The expression $x + 1 - c_1 - c_2x^2$ has, within $[-1, 1]$, derivative equal to 0 at $x = \frac{1}{2}c_2$ if $| c_2 | \geqq \frac{1}{2}$ and is monotone on $\mathfrak{X}$ if $| c_2 | < \frac{1}{2}$. Thus, a routine computation of $\max_x | x + 1 - c_1 - c_2x^2 |$ leads to the conclusion that any c with $c_1 + c_2 = 1$ and $| c_2 | \leqq \frac{1}{2}$ is Chebyshev; i.e., $p = 1$. Hence, $k - p = 2$, and indeed the design ξ^* for which $\xi^*(-1) = \xi^*(1) = \frac{1}{2}$ is optimum. The heart of the matter is that $(1, x^2)$ is not a Chebyshev system and that is is possible to estimate a_3 optimally without estimating a_2 at all.

Example 3.3. *An example where the optimum* ξ^* *is not unique*. There are many obvious examples of this kind, as we have indicated in the paragraph following the proof of Theorem 3. For example, one simple example is given by $\mathfrak{X} = [-1, 1]$, $k = 2$, $f_1(x) = 1, f_2(x) = 1 + \sin 10\, x$ (any ξ which assigns measure $\frac{1}{2}$ to each of the sets where $\sin 10\, x = 1$ or -1 satisfies (2.10)); an even more trivial one is $k = 1, f_1(x) = 1$, where every strategy is optimum.

Example 3.4. *An example where a nonoptimum* ξ *may be preferable*. This example illustrates the phenomenon alluded to in the text, wherein a design ξ which is not optimum in the sense defined in Section 2 may be preferable to an optimum design ξ^* for use (e.g., in testing a hypothesis about a_2) because the latter yields one less degree of freedom for the estimate of σ^2. Let ϵ be a fixed small positive number, and suppose that $\mathfrak{X}$ consists of the three integers 0, 1, and 2, that $k = 2$, and that $f_1(x) = x^2$ and $f_2(x) = 1 + (1 + \epsilon)x$. It is easily computed that the Chebyshev c^* is $1 + 3\epsilon/5$, that $B(c^*)$ consists of the points 1 and 2, that $m(c^*) = 1 + 2\epsilon/5$, and that the optimum ξ^* is given by $\xi^*(1) = 1 - \xi^*(2) = 4/5$. Thus, the efficiency of the design which takes all observations at $x = 0(\xi(0) = 1)$ and estimates a_2 in the obvious way, is $(1 + 2\epsilon/5)^{-2}$; when ϵ is small, this is more than offset by the extra degree of freedom for estimating σ^2 (e.g., 4 for the latter design against 3 for ξ^*, when 5 observations are taken), for the problem of testing a hypothesis about a_2 or giving a confidence interval on a_2.

Example 3.5. *A multidimensional example*. Let $\mathfrak{X}$ be the set of all points (x_1, x_2) in the Euclidean plane for which $| x_1 | \leqq 1$ and $| x_2 | \leqq 1$. Let $k = 6$ and suppose that the functions f_i are, in order, $1, x_1, x_2, x_1^2, x_2^2$, and x_1x_2; thus, for example, we may be testing the hypothesis that a quadratic function of two variables has no interaction term $a_6x_1x_2$, i.e., that $a_6 = 0$. An easy approach to obtaining an optimum ξ is the third method mentioned in Section 2: An obvious guess of a ξ which might be optimum is that measure ξ' (say) which assigns probability $\frac{1}{4}$ to each corner of the square $\mathfrak{X}$. Thus, writing $c_1 + c_4 + c_5 = \bar{c}$, we see that $K(\xi', c)$ is symmetric in each of the variables c_2, c_3, and $\bar{c}$ (which are the only quantities on which it depends), so that $\min_c K(\xi', c) = K(\xi', c') = 1$ is attained for any c' for which $c_2 = c_3 = \bar{c} = 0$. Let c'' have all five of its components equal to zero. Then, clearly, $\max_x K(x, c'') = 1$. Thus, by the discussion following Theorem 2, we have proved that ξ' is optimum. Another way of verifying the optimality of

ξ' is to note that, in the terminology of the remark on symmetry of Section 2, G is the group of symmetries of the square, and an analogue of the last argument mentioned there for the case of polynomial regression with h odd obtains ξ' from the optimum design τ^* which assigns mass 1 to $(1, 1)$ for the problem of estimating a_6 on $0 \leqq x \leqq y \leqq 1$ when the regression function is $a_6 xy$. We note that only a_2, a_3, a_6, and $a_1 + a_4 + a_5$ are estimable for this design. The fact that only four linearly independent estimable linear parametric functions exist here is reflected in the fact that, in the notation of Section 2, $p = 2$. This can be seen by noting that, if $c' = (\epsilon + \delta, 0, 0, -\epsilon, -\delta)$, where ϵ and δ are sufficiently small, then $\max_x K(x, c')$ is still equal to unity, so c' is Chebyshev.

Other examples. Many other examples of optimum designs can be obtained from the extensive literature on Chebyshev approximation problems. For example, Section 37 of [2] can be used to obtain such a design for the setup of Example 3.1 wherein f_k is altered to $f_k(x) = 1/(x - c)$ with $c > b$.

4. The case of several regression coefficients. We consider now the setup of (2.1)–(2.3) (see also (2.18)) in the case where we are interested in inference about more than one of the a_i. In some estimation problems, a treatment like that of Section 5, wherein the behavior of the function $\sum a_i f_i$ rather than that of the a_i themselves is considered, will seem appropriate. However, in most problems of testing hypotheses, as well as in many problems of estimation (especially where the inference is not about all of the a_i), the treatment of the present section may seem appropriate.

We must first choose a criterion of optimality of a design for a problem of estimation or testing hypotheses about s of the a_i, say $a_{k-s+1}, \cdots, a_k$. Of course, it is easy to specify a loss function and a criterion (minimax, etc.) for choosing a design and associated decision procedure; but, as shown in [1], such a simple criterion as that of maximizing the minimum power of a test on an appropriate contour (M-optimality) will usually lead to most unwieldy computations. Even the corresponding local criterion on the power near the null hypothesis (L-optimality) will lead to difficult computations. Two other criteria considered in [1] are D-optimality and E-optimality. In the present setting, n being fixed, a design d^* is said to be D-optimum if $a_{k-s+1}, \cdots, a_k$ are all estimable under d^* and if, among all designs for which these parameters are estimable, denoting by $\sigma^2 V_d$ the covariance matrix of the b.l.e.'s of these parameters when design d is used, det V_d is a minimum for $d = d^*$. A design is said to be E-optimum in the above setting if the maximum eigenvalue of V_d is a minimum for $d = d^*$. The relevance of these criteria for problems of testing hypotheses and of estimation was indicated in [1] and the reference cited there. It was shown that D-optimality is generally more meaningful. There is an additional reason why this is so in problems of the type considered here: Consider the polynomial setup of Example 3.1 for any value $k > 2 (h > 1)$ and $s > 1$. It is clear that the change of scale $x' = hx$ does *not* leave invariant the criterion of E-optimality: a change in the scale of measurement can change the E-optimum design. This is unsatisfactory from both an intuitive point of view (the optimum design depends on the choice of a unit of scale) and from a practical one; one would have to table optimum designs in such problems, as a function of α, β. (A similar remark, of course, applies to L-optimum and M-optimum designs.) On the other hand, D-optimality is invariant under such transformations. The same result is true under a change of origin (or a change of both scale and origin) in this polynomial example: D-optimality is invariant, but E-optimality is not.

Thus, although D-optimality is not an appropriate criterion in all problems, for the reasons given in the previous paragraph it seems reasonable to investigate this criterion as a first attack on the problem of finding optimum designs. We shall thus develop a method for obtaining D-optimum designs in the remainder of this section, except that we shall indicate briefly at the end of this section how various other criteria can be treated similarly.

Proceeding as in Section 2, let h_{tj} be numbers such that, for $i \leqq k - s < t$, the functions f_i are orthogonal to the functions $f_t^* = f_t - \sum_{j=1}^{k-s} h_{tj} f_j$ in the sense of (2.5), i.e.,

$$\sum_{r=1}^{n} f_i(x_n) f_t^*(x_r) = 0, \qquad i \leqq k - s < t. \tag{4.1}$$

Then, as in the discussion of (2.6), we see that σ^2 times the inverse of the covariance matrix $\sigma^2 V_d$ of best linear estimators of $a_{k-s+1}, \cdots, a_k$ has elements $\sum_r f_i^*(x_r) f_j^*(x_r)$, $k - s < i, j \leqq k$. For $t > k - s$, let $f_t^{**} = f_t^* - \sum_{j<t} g_{tj} f_j^*$ be orthogonal to f_j^* for $k - s < j < t$. Since the linear transformation which takes the f_t^* into the f_t^{**}, $k - s < t \leqq k$, has determinant 1, and since $\sum_r f_i^{**}(x_r) f_j^{**}(x_r) = 0$ if $k - s < i < j$, we obtain

$$\det V_d^{-1} = \prod_{i>k-s} \sum_r [f_i^{**}(x_r)]^2. \tag{4.2}$$

Now, f_i^{**} is clearly f_i minus the projection of f_i on the linear space spanned by $f_1, f_2, \cdots, f_{i-1}$. Thus, the ith term in the product of (4.2) is just the expression of (2.7) with k replaced by i. Finally, then, making the same approximation as in Section 2 regarding the representation of the class of all designs by the class of all probability measures ξ on $\mathfrak{X}$, we have demonstrated, to within this approximation, the validity of the following definition, wherein $c^{(j)}$ denotes a vector $(c_1^{(j)}, \cdots, c_{j-1}^{(j)})$ of $j - 1$ components:

Definition. *A measure ξ^* in Ξ is said to be D-optimum (for the parameters $a_{k-s+1}, \cdots, a_k$) if*

$$\begin{aligned}\prod_{j>k-s} \min_{c^{(j)}} \int [f_j(x) - \textstyle\sum_1^{j-1} c_i^{(j)} f_i(x)]^2 \xi^* (dx) \\ = \max_{\xi \epsilon \Xi} \prod_{j>k-s} \min_{c^{(j)}} \int [f_j(x) - \sum_1^{j-1} c_i^{(j)} f_i(x)]^2 \xi (dx)\end{aligned} \tag{4.3}$$

Of course, (4.3) reduces to (2.8) in the case $s = 1$. When f_1 is a constant, a ξ which is optimum for $s = k - 1$ is also optimum for $s = k$.

We note that it is a consequence of this definition that an optimum design is optimum for all values of σ^2.

For the special case where $s = k$ and $\mathfrak{X}$ consists of k points, it is easy to prove that the unique optimum ξ puts mass $1/k$ on each point. For if A is the matrix whose (i, j) element is $f_i(x_j)$ and B is the diagonal matrix with $\xi(x_j)$ the diagonal element in the jth row, an optimum design maximizes $\det(ABA') = (\det A)^2 \det B$. This argument has been employed by Hoel in the problem considered by him; see Example 4 below.

The methods of Section 2 do not directly yield anything here for the general problem. The analogue of Theorem 1 is essentially empty, since the various $B(c^{(j)})$'s for $c^{(j)}$ Chebyshev will not in general coincide. The game-theoretic approach is inapplicable because the product on the left side of (4.3) is not linear in ξ^*; moreover, the product of the integrals (before minimizing over the $c^{(j)}$) is not convex in the $c^{(j)}$'s, since u^2v^2 is not a convex function of u and v. The following analysis will, however, yield a method for obtaining an optimum ξ.

For $j > k - s$, let

$$F_j(\xi) = \min_{c^{(j)}} \int [f_j(x) - \sum_1^{j-1} c_i^{(j)} f_i(x)]^2 \xi\,(dx). \tag{4.4}$$

In s-dimensional Euclidean space R^s, let S be the set of all points $F(\xi) = (F_{k-s+1}(\xi), \cdots, F_k(\xi))$ for ξ in Ξ. Although S may not be convex, it possesses the following "upper convexity" property, which is all we require: For any ξ_1 and ξ_2 in Ξ and any λ with $0 < \lambda < 1$,

$$F_j(\lambda\xi_1 + (1 - \lambda)\xi_2) \geqq \lambda F_j(\xi_1) + (1 - \lambda)F_j(\xi_2) \tag{4.5}$$

for all $j > k - s$. In fact, (4.5) is an immediate consequence of the linearity in ξ of the integral of (4.4).

Let $u_{k-s+1}, \cdots, u_k$ be the coordinate functions of R^s. For $\delta > 0$, let G_δ be the set of all points in R^s with all coordinates positive and $\prod_j u_j \geqq \delta$. Let G_δ' be the subset of G_δ where $\prod_j u_j = \delta$. We note that G_δ is convex. Suppose that S is closed (this is easily proved from (4.4) if $\mathcal{B}$ is large enough so that Ξ is compact; the modification which is needed if Ξ is not closed is trivial, anyway), and let δ_0 be the largest value of δ such that G_δ and S have a nonempty intersection. (Such a δ_0 exists since S has points with all coordinates positive.) If T is the convex hull of S, property (4.5) implies that δ_0 is also the largest value of δ such that G_δ and T have a nonempty intersection. Hence, applying the separation theorem for G_{δ_0} and T, we conclude that there is a hyperplane L with positive direction cosines such that L separates G_{δ_0} and S. Thus, any point $F(\xi^*)$ in $G_{\delta_0} \cap S$ clearly maximizes $\prod_j F_j(\xi)$ (i.e., that ξ^* satisfies (4.3)); and, for positive numbers λ_j' for which L is given by $\sum \lambda_j' U_j =$ constant, that point maximizes $\sum_j \lambda_j' F_j(\xi)$. Finally, since all points of G_σ' are extreme, L intersects G_{δ_0} in exactly one point, as does therefore S.

Before summarizing the above results, we note that, for $\lambda = (\lambda_{k-s+1}, \cdots, \lambda_k)$

with all $\lambda_i > 0$, the payoff function

$$K_\lambda(x, c) = \sum_{j>k-s} \lambda_j [f_j(x) - \sum_{i<j} c_i^{(j)} f_i(x)]^2, \tag{4.6}$$

where $c = (c^{(k-s+1)}, \cdots, c^{(k)})$, satisfies all of those conditions satisfied by the function K of Section 2 which were used in the proof of the game-theoretic results of the lemma there. Thus, that lemma is valid when K is replaced by K_λ.[4] The function K_λ is of course no longer in a form suitable to make use of Chebyshev approximation results. However, for any λ, if c_λ^* is minimax for the payoff function K_λ, we can still characterize maximin ξ_λ's in terms of the set $\bar{B}_\lambda(c_\lambda^*)$ (say), defined to be the set of x for which $K_\lambda(x, c_\lambda^*)$ achieves its maximum. With this interpretation of symbols, the analogue of (2.10) is proved here exactly as in (2.11).

We have thus proved that following,[5] where C now stands for the set of vectors $c = (c^{(k-s+1)}), \cdots, c^{(k)})$ and $c_\lambda^* = \{c_i^{(j)*}\}$ stands for a vector of this type:

THEOREM 4. *The game of Ξ vs. C with payoff function K_λ is determined. If ξ_λ is maximal with respect to c_λ^* while c_λ^* is minimal with respect to ξ_λ, then ξ_λ is maximin. Thus, if c_λ^* is minimax and*

$$\xi_\lambda(\bar{B}_\lambda(c_\lambda^*)) = 1 \tag{4.8}$$

and

$$\int [f_j(x) - \sum_{i<j} c_i^{(j)*} f_i(x)] f_i(x) \xi_\lambda(dx) = 0 \tag{4.9}$$

for $i < j$ and $k - s < j \leqq k$, then ξ_λ is maximin; moreover, every maximin ξ_λ satisfies (4.8) and (4.9) for every minimax c_λ^*. There is, to within a multiplicative constant, a unique value λ^* of λ such that $\prod_i F_i(\xi_\lambda)$ is a maximum for $\lambda = \lambda^*$ and some ξ_{λ^*}. Those ξ_{λ^*} which maximize $\prod_i F_i(\xi_{\lambda^*})$, and no other ξ's, are optimum. $F(\xi_{\lambda^*})$ is the same for any optimum ξ_{λ^*}.

We now consider an example.

Example 4. Consider the setup of Example 3.1, where (see the end of the second paragraph of the present section) we may suppose $\alpha = -1$, $\beta = 1$. Suppose $k = 3$ ($h = 2$), and $s = 2$; as we have remarked earlier, the optimum design obtained below will also obviously be D-optimum for the case $s = 3$. An

[4] That part ot the lemma which concerns the number $k - p$ is valid when k is replaced by $1 + s(2k - s - 1)/2$ ($= 1 +$ number of components of c) in the statement of the lemma. However, this is of no use to us since it may be that no maximin strategy on the specified number of points is optimum. For example, in the set-up of Example 5.2 below with $s = k = 2$, one can verify that the λ^* of Theorem 4 is (15/4, 1), and that any ξ_λ^* with first and second components equal is maximin, but only (4/15, 4/15, 7/15) is optimum.

It is trivial that the optimum strategy need be concentrated on no more than $1 + k(k + 1)/2$ points. For the criterion of optimality (4.3) involves ξ only through the elements (5.2) below of the matrix $M(\xi)$. These matrices form a convex body of dimensionality at most $k(k + 1)/2$, spanned by matrices of ξ's concentrated on a single point Hence any $M(\xi)$ is a linear convex combination of at most $1 + k(k + 1)/2$ extreme elements.

[5] See also footnote 6.

elegant solution to this problem for general k and $s = k$, has been given by P. G. Hoel [15] (see also Example 5.1 below). The case $s < k - 1$ does not seem to yield to his attack. The present problem is discussed here as an illustration of our methods. We may take 1 and γ for the components of λ, and write $K'_\gamma(x, d) = (x - d'_0)^2 + \gamma(x^2 - d''_1 x - d''_0)^2$ in place of K_λ. For fixed γ, one may *guess* that there will be a maximin strategy ξ'_γ of the form $\xi'_\gamma(-1) = \xi'_\gamma(1) = \alpha_\gamma$, $\xi'_\gamma(0) = 1 - 2\alpha_\gamma$, for some α_γ. With respect to such a ξ_γ, the minimal strategy (which must merely satisfy the orthogonality relation (4.9)) is obviously $d'_0 = d''_1 = 0$, $d''_0 = 2\alpha_\gamma$. For this choice d_γ (say) of d, we obtain $K'_\gamma(\xi'_\gamma, d_\gamma) = \gamma[2\alpha_\gamma - 4\alpha_\gamma^2] + 2\alpha_\gamma$. This is maximized by $a_\gamma = \min(\frac{1}{2}, (\gamma + 1)/4\gamma)$, and for the strategy $\xi_{\gamma*}$ corresponding to this value of α_γ we obtain

$$\min_d K'_\gamma(\xi^*_\gamma, d) = \begin{cases} (\gamma + 1)^2/4\gamma & \text{if } \gamma > 1, \\ 1, & \text{if } \gamma \leqq 1. \end{cases} \tag{4.10}$$

On the other hand,

$$\min_d \max_x K'_\gamma(x, d) \leqq \max_x K'_\gamma(x, d_\gamma). \tag{4.11}$$

Since $K'_\gamma(x, d_\gamma)$ is convex in x^2, its maximum is attained at either $x^2 = 0$ or $x^2 = 1$, and an easy computation shows that the right side of (4.11) is in fact equal to the right side of (4.10). Thus, we have proved that ξ^*_γ is maximin. Finally, $F_2(\xi^*_\gamma)F_3(\xi^*_\gamma) = 4\alpha_\gamma^2(1 - 2\alpha_\gamma)$, which is maximized by $\alpha_\gamma = \frac{1}{3}$. Thus, an optimum design for this problem is $\xi(-1) = \xi(0) = \xi(1) = \frac{1}{3}$. Of course the optimum designs for a given set of f_i will depend on s, as exemplified by the different results obtained in Example 3.1 and Example 4.

We shall now mention briefly methods for obtaining designs which are optimum in two other senses. Although it is not difficult to characterize E-optimum procedures in *simple* examples, they often seem much harder to calculate than D-optimum ones. Somewhat easier is the characterization of that design which minimizes the maximum eigenvalue of the covariance matrix of best linear estimators of the regression coefficients of the f_t^{**} (the regression function being expressed in terms of the f_t for $t \leqq k - s$ and of the f_t^{**} for $t > k - s$); i.e., of $L_d V_d L'_d$ where L_d is a square matrix with ones on the main diagonal and zeros above it. (The f_t^{**} depend on the design, which indicates the intuitive weakness of this criterion; however, as pointed out in [1], the criterion of E-optimality, which has often been considered in the literature, suffers from a similar shortcoming.) Again making the approximation that we do not restrict $n\xi$ to be integer-valued, this criterion amounts to finding that ξ which maximizes $\min_{j>k-s} F_j(\xi)$, i.e., if δ' is the largest value of δ for which the orthant $H_\delta = \{\min_j u_j \geqq \delta\}$ intersects S nonvacuously, those ξ for which $F(\xi)$ is in $H_{\delta'} \cap S$ are the optimum procedures with respect to this criterion. Another criterion which has been considered in the literature, especially in estimation problems, is that of minimizing the "average variance", $\sigma^2 s^{-1}$ trace (V_d). Defining $F^*_j(\xi')$ to be the expression of (4.4) with the sum in the integrand taken only from 1 to $k - s$, this criterion amounts to minimizing $\sum_{j>k-s} F_j^*(\xi)$. Replacing S by the set of points $F^*(\xi) = (F_{k-s+1}^*(\xi), \cdots, F_k^*(\xi))$, and restricting the sum over i in (4.6) to values $\leq k - s$, this amounts to finding the maximin ξ's for a λ with all components equal. These maximin ξ's for the *original* S and K_λ (with all λ_j equal) would of course minimize the average variance of the b.l.e.'s of regression coefficients of the f_i^{**}; i.e., would minimize the trace of $L_d V_d L_d'$. Criteria like that of minimizing the average variance are subject to the same criticisms as E-optimality.

Remarks. As in the problem of Section 2, one can prove that the symmetry condition (2.15) and the obvious analogue of the condition of the line above (2.15) imply the existence of a symmetrical optimum ξ for any of the criteria considered in the present section. For example, from (4.5) it follows at once that, if ξ is D-optimum, then the symmetrical $\bar{\xi}$ defined by (2.17) is also D-optimum. Remarks analogous to those of Section 2 on the number of points at which a symmetrical optimum ξ will be concentrated, clearly hold in the problems of this section. We note that the choice of the form of ξ_γ' in Example 4 is motivated by symmetry considerations, although the optimum weights must be computed in any approach.

The remark concerning the modification of (2.18) applies also to the problems of this section.

5. Estimation of the whole regression function. In the setup described by (2.1–(2.3), suppose the problem is one of estimation concerning all the a_i. One approach has been indicated in Section 4. Another approach is to think of the problem not as one of estimating the parameters a_i, but rather as one of estimating the entire function $\sum a_i f_i$. Thus, if g is the estimate of $\sum a_i f_i$, it is desired to make some measure of the average deviation of g from $\sum a_i f_i$ small in some sense, by choosing an appropriate design. The most obvious possibilities of such measures are perhaps (1) $\sup_a EW(\sup_x | g(x) - \sum a_i f_i(x) |)$, where W is nondecreasing; (2) the integral with respect to some measure μ on $\mathfrak{X}$ of $\sup_a EW(| g(x) - \sum a_i f_i(x) |)$; (3) the supremum on $\mathfrak{X}$ and $\mathfrak{A}$ of $EW(| g(x) - \sum a_i f_i(x) |)$. Of these three possibilities, the first is perhaps the most meaningful for most applications (with perhaps the inclusion of a weight function $h(x)$ multiplying $| g(x) - \sum a_i f_i(x) |$) but is computationally much more difficult to treat than the others; the second possibility is by far the easiest computationally, but is least satisfactory from a practical point of view because of the necessity of choosing μ—for example, if $\mathfrak{X}$ is a line segment, the optimum design will not be invariant under homeomorphisms of $\mathfrak{X}$, if μ is always chosen to be Lebesgue measure; the third possibility is a compromise between the first two and, as a first attack on the problem, is what we consider in this section, with $W(t) = t^2$. We note that a remark of [9, p. 215] indicates that Box and Hunter are considering the second approach for certain polynomial multiple regression problems when $W(t) = t^2$ and μ is Lebesgue measure on a Euclidean set. We note that it is a consequence of all three definitions of optimality discussed in this paragraph that an optimum design is optimum for all values of σ^2.

We shall not have to concern ourselves here with the choice of the function g: for example, the remarks of Section 2 extend here to show that, for a given design, if $\hat{a}$ is the b.l.e. of a, then $\sup_{a,x} E_a[\sum a_i f_i(x) - g(x)]^2$ is a minimum for $g(x) = \sum \hat{a}_i f_i(x)$. We therefore assume this choice of g in what follows. Thus, we are led to consider the minimization with respect to the design d of the expression

$$(5.1) \qquad \max_x E[\textstyle\sum_i(\hat{a}_i - a_i)f_i(x)]^2 = \sigma^2 \max_x f(x)'V_d f(x),$$

where we have written $f(x)$ for the vector of $f_i(x)$'s. Using again the representation of a design as a measure ξ, the analogue of V_d is the inverse of the matrix $M(\xi)$ whose (i, j)th element is

$$(5.2) \qquad m_{ij}(\xi) = \int f_i(x)f_j(x)\xi(dx).$$

Thus, making an approximation like that of Section 2 in not requiring $n\xi$ to be integral, we define a design ξ^* to be optimum for the problem of this section if $M(\xi^*)$ is nonsingular and

$$(5.3) \qquad \max_x f(x)'M(\xi^*)^{-1}f(x) = \min_{\xi\epsilon\Xi} \max_x f(x)'M(\xi)^{-1}f(x).$$

It seems more difficult here than in Section 2 to give a useful general computing algorithm. We now describe one device which seems useful in many examples. Let D_ξ be a non-singular matrix such that the vector $g_\xi = D_\xi f$ consists of functions $g_{\xi,i}$ which are orthonormal with respect to ξ; it is clear that such a D_ξ exists for any ξ in Ξ for which $M(\xi)$ is non-singular. Since the (i, j)th element of $D_\xi M(\xi)D_\xi'$ is the integral with respect to ξ of $g_{\xi,i}g_{\xi,j}$, we obtain

$$(5.4) \qquad f(x)'M(\xi)^{-1}f(x) = g_\xi'(x)(D_\xi M(\xi)D_\xi')^{-1}g_\xi(x) = \textstyle\sum_i [g_{\xi,i}(x)]^2$$

(Since the left side of (5.4) does not depend on D_ξ, neither does the right side; thus, in searching for a ξ which minimizes the maximum with respect to x of (5.4), it suffices to consider for each ξ only that D_ξ and g_ξ which are computationally most convenient.) Since the $g_{\xi,i}$'s are orthonormal with respect to ξ, the integral with respect to ξ of the last expression of (5.4) is k, and this cannot be greater than the maximum with respect to x of (5.4). Thus, a sufficient condition for a given ξ to be an optimum design is

$$(5.5) \qquad \max_x \sum_{i=1}^{k} [g_{\xi,i}(x)]^2 = k.$$

Of course, a necessary condition for (5.5) to be satisfied is that ξ give measure one to the set of x where $\sum[g_{\xi,i}(x)]^2 = k$, and it is useful to keep this in mind in examples.

Suppose (5.5) is satisfied for a ξ' concentrated on k points, say $x_1, \cdots, x_k$. Then the $k \times k$ matrix whose (i, j)th element is $g_{\xi',i}(x_j)[\xi'(x_j)]^{\frac{1}{2}}$ has orthonormal rows and, hence, orthonormal columns: $\sum_i [g_{\xi',i}(x_j)]^2 \xi_j' = 1$ for $1 \leq j \leq k$.

Hence, $\xi_j' > 0$, and by (5.5) $\xi_j' \geqq 1/k$. Hence, each ξ_j is $1/k$. We summarize our results.

THEOREM 5. *If* (5.5) *holds, then* ξ *is optimum.*[6] *If* (5.5) *holds for a* ξ *concentrated on* k *points, then* ξ *gives measure* $1/k$ *to each of these points.*

If the setup is that of Example 3.1 it follows from the results of [10] cited in Section 2 that there exists an optimum ξ concentrated on exactly k points. This will not be true in general (see Example 5.2 below).

In the special case where $\mathfrak{X}$ consists of k points, the argument of the paragraph preceding the present theorem, applied to the distribution $\xi_j = 1/k, j = 1, \cdots, k$, shows that (5.5) is satisfied for this distribution, and hence the latter is optimum. Combining this with a remark which follows (4.3) we conclude that, when $\mathfrak{X}$ consists of k points, the design which puts mass $1/k$ on each point is the unique optimum design according to both the definition (4.3) for $s = k$ (the problem of Section 4) and the definition (5.3) (the problem of the present section).

Example 5.1. *The setup of Example* 3.1. It is possible to solve this problem by our methods and such a solution was given in the original draft of this paper. In the meantime, however, a solution has been published by Guest [13], so that there is no point to repeating the details of our solution. An earlier discussion by Smith [14] gave details of designs up to $k = 7$. The optimum design assigns mass $1/k$ to the points $+1$, -1, and the roots of $L_h'(x) = 0$, where L_h' is the derivative of the Legendre polynomial. (5.5) is satisfied ([13], equation (10)). It therefore follows from Theorem 6 below that this design is also optimum in the sense of definition (4.3) for $s = k$ (problem of Section 4) for this setup; i.e., a special case of Theorem 6 asserts that Hoel's design [15] is the same as that of Guest [13].[7] This last fact was noted by Hoel through an examination of the explicit results in the polynomial case.

Example 5.2. This example illustrates the use of Theorem 5 where the optimum ξ is concentrated on more than k points and does not give equal measure to all of them. Let $k = 2$ and let $\mathfrak{X}$ consist of three points. Thus, we hereafter write the f_i and ξ and S as triples, where $S(x) = \sum_i [g_{\xi,i}(x)]^2$. Suppose $f_1 = (1, 1, 0)$ and $f_2 = (0, 1, 2)$. For $\xi = (\xi_1, \xi_2, \xi_3)$, we obtain easily

$$S = (\xi_1\xi_2 + 4\xi_2\xi_3 + 4\xi_1\xi_3)^{-1}(\xi_2 + 4\xi_3, \xi_1 + 4\xi_3, 4\xi_1 + 4\xi_2)$$

We have $\sum \xi_i S_i = 2$, identically in ξ. Suppose $\xi_1 = 0$, $\xi_2 > 0$, $\xi_3 > 0$. Then either 1) $S_2 = S_3$, in which case $\xi_2 = \xi_3 = \frac{1}{2}$ and $S_1 > S_2 = 2$, or 2) $\max(S_2, S_3) > 2$. Thus, in either case $\max_i S_i > 2$. A similar argument applies if either of the other ξ_i's is 0, and two ξ_i's can obviously not be 0. Thus, $\max_i S_i$ can be 2 only if all ξ_i are positive and all S_i are equal to 2. The unique optimum ξ is thus easily seen to be (4/15, 4/15, 7/15).

---

[6] The converse of this statement is true. In fact, it will be proved in a subsequent paper (the results were obtained too late for inclusion in the present paper) that the following three statements are equivalent: (a) the design ξ is optimum in the sense of Section 4 with $s = k$; (b) the design ξ is optimum in the sense of Section 5; (c) the design ξ satisfies (5.5).

[7] This is a special case (for polynomial regression) of the result described in footnote 6.

It is obvious how to give examples like those of Section 3 where the optimum ξ is not unique, etc.

The argument just after (2.17) is easily modified to apply to the expression on the left side of (5.1), so that we can again conclude that there exists an optimum symmetrical ξ if (2.15) and the obvious analogue of the condition of the line above (2.15) hold. The question of the number of points at which an optimum symmetrical ξ will be concentrated is difficult, as is the corresponding question for general optimum ξ.

The modification of (2.18) can be made in the problem of this section, exactly as in Section 2.

We shall conclude this section with a result which sheds some light on the connection between the problem of Section 4 for $s = k$ and the problem of the present section. This result has already been cited in Example 5.1.

THEOREM 6.[8] *If the design which puts mass $1/k$ on each of k points satisfies* (5.5) *and is optimum in the sense of* (5.3) (*problem of Section* 5), *then this design is also optimum in the sense of* (4.3) (*problem of Section* 4) *with $s = k$.*

PROOF: Let ξ_0 be a design, optimum in the sense of (5.3), such that ξ_0 assigns mass $1/k$ to each of the points $x_1, \cdots, x_k$ in $\mathfrak{X}$, and such that (5.5) is satisfied for $\xi = \xi_0$. Since a design optimum for the problem of Section 4 with $s = k$ is invariant under a linear transformation on the f_i, it will suffice to prove that ξ_0 is optimum for this problem assuming $f_i = g_{\xi_0, i}$; henceforth we make this assumption. Thus

$$\max_x \sum_i f_i^2(x) = k \tag{5.6}$$

and

$$m_{ij}(\xi_0) = \delta_{ij},$$

and we have to prove that ξ_0 maximizes $\det M(\xi)$. Now from (5.6) for any ξ we have

$$\sum_i m_{ii}(\xi) \leqq k$$

and hence

$$\det M(\xi) \leqq \prod_i m_{ii}(\xi) \leqq 1 = \det M(\xi_0)$$

This proves the theorem.

The authors are obliged to Professor G. Elfving for helpful comments.

REFERENCES

[1] J. KIEFER, "On the nonrandomized optimality and randomized nonoptimality of symmetrical designs", *Ann. Math. Stat.*, Vol. 29 (1958), pp. 675–699.

[2] N. I. ACHIESER, *Theory of Approximation*, Ungar Pub. Co., New York, 1956.

[8] Theorem 6 is a very special case of the results announced in footnote 6.

[3] G. ELFVING, "Optimum allocation in linear regression theory", *Ann. Math. Stat.*, Vol. 23 (1952), pp. 255–262.

[4] G. ELFVING, "Geometric allocation theory", *Skand. Akt.* 1955, pp. 170–190.

[5] H. CHERNOFF, "Locally optimum designs for estimating parameter" *Ann. Math. Stat.*, Vol. 24 (1953), pp. 586–602.

[6] H. F. BOHNENBLUST, S. KARLIN, AND L. SHAPLEY, "Games with continuous, convex payoff", *Ann. Math. Studies*, No. 24, pp. 181–192.

[7] J. KIEFER, "Invariance, minimax sequential estimation, and continuous time processes", *Ann. Math. Stat.*, Vol. 28 (1957), pp. 573–601.

[8] R. A. FISHER, *Statistical Methods for Research Workers*, tenth edition, Oliver and Boyd, Edinburg, 1946.

[9] G. E. P. BOX AND J. S. HUNTER, "Multi-factor experimental designs for exploring response surfaces", *Ann. Math. Stat.* Vol. 28 (1957), pp. 195–241.

[10] J. A. SHOHART AND J. D. TAMARKIN, *The Problems of Moments*, *Math. Surveys*, No. 1, Amer. Math. Soc., New York, 1943.

[11] E. J. WILLIAMS, "Optimum allocation for estimation of polynomial regression," (abstract) *Biometrics*, Vol. 14 (1958), p. 573.

[12] S. EHRENFELD, "Complete class theorems in experimental design", *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1955.

[13] P. G. GUEST, "The spacing of observations in polynomial regression", *Ann. Math. Stat.*, Vol. 29 (1958), pp. 294–299.

[14] K. SMITH, "On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations", *Biometrika*, Vol. 12 (1918), pp. 1–85.

[15] P. G. HOEL, "Efficiency problems in polynomial estimation", *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1134–46.