OPTIMUM DESIGNS IN REGRESSION PROBLEMS, II

BY J. KIEFER[1]

Cornell University

0. Summary. Extending the results of Kiefer and Wolfowitz [10], [11], methods are obtained for characterizing and computing optimum regression designs in various settings, and examples are given where D-optimum designs are computed.

In Section 1 we introduce the main definitions and notation which will be used in the paper, and discuss briefly the roles of invariance, randomization, number of points at which observations are taken, and nonlinearity of the model, in our results.

In Section 2 we prove the main theoretical results. We are concerned with the estimation of s out of the k parameters, extending an approach developed in [10] and [11] in the case $s = k$. There is no direct way of ascertaining whether or not a given design ξ^* is D-optimum for (minimizes the generalized variance of the best linear estimators of) the s chosen parameters, and Theorems 1 and 2 provide algorithms for determining whether or not a given ξ^* is D-optimum. If all k parameters are estimable under ξ^*, we can use (2.7) to decide whether ξ^* is D-optimum, while if not all k parameters are estimable we must use the somewhat more complicated condition (2.17) (of which part (a) or (b) is necessary for optimality, while (a), (c), or (d) is sufficient). An addition to Theorem 2 near the end of Section 3 provides assistance in using (2.17) (b). Theorem 3 of Section 2 characterizes the set of information matrices of the D-optimum designs.

In Section 3 we give a geometric interpretation of the results of Section 2, and compare the present approach with that of [10]. In the case $s = k$, the present approach reduces to that of Section 5 of [10] and of [11]. When $1 < s < k$, we obtain an algorithm which differs from that of Section 4 of [10] and which appears to be computationally easier to use. When $s = 1$, the results of the present paper are shown to reduce to those of Section 2 of [10]; in particular, we obtain the game-theoretic results without using the game-theoretic machinery of [10].

In Section 4 we determine D-optimum designs for the problems of quadratic regression on a q-cube and polynomial regression on a real interval with $1 < s < k$.

Part II of the paper is devoted entirely to the determination of D-optimum designs for various problems in the setting of simplex designs considered by Scheffé [12].

Received December 28, 1959; revised October 12, 1960.

[1] Research sponsored by the Office of Naval Research and the Army Signal Corps.

Various unsolved problems are mentioned throughout the paper. Further examples will be published elsewhere.[2]

PART I. GENERAL CONSIDERATIONS

1. Introduction. The design of optimum experiments in regression settings can be a tedious computational problem. In the present paper we are concerned with the development and application of algorithms which make this task easier. Early work in this area was done by Elfving [4], Chernoff [1], and Ehrenfeld [2]. The characterization of optimum designs in general symmetrical settings where block designs are customarily employed was given by the present author in [8]. As described in Section 0, the present paper continues the work of Kiefer and Wolfowitz [10], [11].

We now introduce our terminology and notation, which is essentially that of [9] and [11]. Let $f_1, f_2, \cdots, f_k$ be given functions on a space $\mathfrak{X}$. To avoid trivial circumlocutions (see [11]), we assume $\mathfrak{X}$ compact and the f_i linearly independent and continuous. By $f(x)$ we denote the column k-vector with components $f_i(x)$. For any discrete probability measure ξ on $\mathfrak{X}$, write

$$m_{ij}(\xi) = \int_{\mathfrak{X}} f_i(x) f_j(x) \xi(dx). \tag{1.1}$$

(As discussed in [10] and [11], other probability measures can be considered, but are not needed.) Write $M(\xi)$ for the $k \times k$ matrix $\{m_{ij}(\xi)\}$. Any such ξ is called an *experiment*, and the set of all ξ will be denoted by Ξ.

The practical meaning of these notions is this: We are concerned with inference regarding an unknown k-vector θ, an element of a k-dimensional Euclidean space Ω. A single observation at the point x (value of the independent variable) in $\mathfrak{X}$ yields a random variable Y_x for which

$$\begin{aligned} EY_x &= \theta' f(x) = \sum_1^k \theta_i f_i(x), \\ \operatorname{Var}(Y_x) &= \sigma^2. \end{aligned} \tag{1.2}$$

Thus, $\theta' f(x)$ is the regression function. If $\eta(x)$ observations out of the total of n uncorrelated observations available in a given experiment are taken at the point x and we let $\xi = n^{-1}\eta$, we obtain $n\sigma^{-2}M(\xi)$ for the "information matrix" of the experiment; thus, for example, if all components of θ are estimable, the covariance matrix of the best linear estimators (b.l.e.'s) of components of θ is $n^{-1}\sigma^2 M^{-1}(\xi)$. The justification for considering only b.l.e.'s in the sequel (whether or not σ^2 is known), and the trivial modifications which are needed in our developments if the Y_x are not uncorrelated with equal variances or do not cost the same amounts, are discussed in [9] and [10]. Also discussed there is the relevance of considering designs ξ which take on values other than multiples of $1/n$. Briefly, such considerations allow us to develop useful computational

[2] Optimum designs for certain problems in the settings where systematic designs and rotatable designs are employed, will appear in the *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*.

techniques which are completely absent if we restrict the values of ξ, and they allow us to find one optimum ξ (rather than a different one for each n) which can immediately be altered into an *actual* design (i.e., a ξ with restricted values) which is within $O(n^{-1})$ of being optimum. Thus, we shall always consider, without restriction, the whole space Ξ of designs.

Suppose we are interested in inference regarding s independent linear parametric functions of θ, which we can take to be $\theta_1, \theta_2, \cdots, \theta_s$ without loss of generality. We partition $M(\xi)$ and $M^{-1}(\xi)$ as

$$\begin{Vmatrix} M_1(\xi) & M_2(\xi) \\ M_2'(\xi) & M_3(\xi) \end{Vmatrix} \quad \text{and} \quad \begin{Vmatrix} M^{(1)}(\xi) & M^{(2)}(\xi) \\ M^{(2)\prime}(\xi) & M^{(3)}(\xi) \end{Vmatrix},$$

respectively, where if $M(\xi)$ is singular we take $M^{-1}(\xi)$ to be a pseudo-inverse (see, e.g., [1] and the next section for details). Here $M_1(\xi)$ and $M^{(1)}(\xi)$ are $s \times s$, and $n^{-1}\sigma^2 M^{(1)}(\xi)$ is the nonsingular covariance matrix of b.l.e.'s of $\theta_1, \cdots, \theta_s$ if all of these parameters are estimable when the design ξ is used. We shall say that ξ^* *is D-optimum for* $\theta_1, \cdots, \theta_s$ if

$$\det M^{(1)}(\xi^*) = \min_{\xi \in \Xi} \det M^{(1)}(\xi), \tag{1.3}$$

i.e., if ξ^* minimizes the generalized variance of the b.l.e.'s of $\theta_1, \cdots, \theta_s$. Extensive discussions in [8], [9], and [10] are concerned with the meaningfulness of this criterion, with certain intuitively appealing properties of the criterion (e.g., invariance under certain transformations), and with a comparison of this criterion with certain other optimality criteria. In the present paper we shall be entirely concerned with D-optimality and an equivalent criterion which is discussed just after (1.4) below; results concerning other optimality criteria are contained in [1], [2], [4], [8], [9], [10], and in the references cited there.

We shall also partition θ (resp., $f(x)$) into $\theta^{(1)}$ and $\theta^{(2)}$ (resp., $f^{(1)}(x)$ and $f^{(2)}(x)$), where $\theta^{(1)}$ (resp., $f^{(1)}(x)$) is an s-vector.

When $s = k$, it will sometimes be convenient to state a problem in terms of the k-dimensional vector space F spanned by the f_i; clearly, D-optimality depends only on F and not on the choice of the f_i used to represent it (analogous remarks apply when $s < k$).

The variance of the b.l.e. of $\theta' f(x)$, the true regression function evaluated at the point x, when the design ξ (for which all components of θ are estimable) is used, is $\sigma^2 n^{-1} d(x, \xi)$, where

$$d(x, \xi) = f(x)' M^{-1}(\xi) f(x). \tag{1.4}$$

Another possible optimality criterion when $s = k$ is for ξ to minimize $\max_x d(x, \xi)$. One of the results of [11] is that this criterion and D-optimality are equivalent, and this fact proves very useful in constructing optimum designs. Thus, the direct maximization of det $M(\xi)$ over *all* ξ will usually be very difficult. However, as in the first example of Section 4, one can often guess a simple (finite-dimensional) subclass of Ξ, compute the ξ^* which maximizes det $M(\xi)$ over this subclass, and then compute $\max_x d(x, \xi^*)$. If this last quantity equals k (and only then), ξ^* is indeed optimum among *all* members of Ξ, as was proved in [10] and [11]. Thus, it would appear useful to find a generalization of $d(x, \xi)$ and the criterion in terms of it, when $s < k$, and to prove the equivalence of this criterion to D-optimality, so as to yield a computational technique in the case $s < k$ which is analogous to that just discussed for the case $s = k$. (Since the matrix $M^*(\xi)$ of (1.5) is not linear in ξ if $s < k$, and since $M(\xi^*)$ can be singular, there will be complications which did not arise in the case $s = k$.) This is the task of Section 2. The method of constructing D-optimum designs developed in this paper is thus to guess a ξ^* and to compute $\max_x d(x, \xi^*)$ (with the definition of (2.3) in place of (1.4)) or, if not all components of θ are estimable, $\max_\xi D(\xi, \xi^*)$. Paralleling the procedure outlined above in the case $s = k$, we then use (2.7) or (2.17) (a) to test the optimality of ξ^*.

Before turning to Section 2, we shall mention a few results which are relevant for the remainder of the paper.

Invariance. If A and B are $s \times s$ nonnegative definite symmetric matrices, write $A \geqq B$ if $A - B$ is nonnegative definite. Without recourse to a specific optimality criterion such as D-optimality, one can study complete classes of designs, admissible designs, etc., and this was first done in the regression setting for $s = k$ by Ehrenfeld [3], who exploited the usefulness of $M(\xi_1) \geqq M(\xi_2)$ as a criterion equivalent to "ξ_1 is at least as good as ξ_2 for all linear estimation problems". As discussed by the present author [9], the idea can also be explained as the sufficiency (in the sense of Blackwell) in the normal case of ξ_1 for ξ_2, criteria other than Ehrenfeld's for completeness can be given, and the whole theory can be extended to the case where we are only interested in s out of the k parameters. In this case, ξ_1 is at least as good as ξ_2 if and only if $M^{(1)}(\xi_1) \leqq M^{(1)}(\xi_2)$ or, equivalently, if $M^*(\xi_1) \geqq M^*(\xi_2)$, where

$$(1.5) \qquad M^*(\xi) = [M^{(1)}(\xi)]^{-1} = M_1(\xi) - M_2(\xi)M_3^{-1}(\xi)M_2'(\xi).$$

(The modification in the singular case is obvious.) If A is a nonsingular $k \times k$ matrix of the form

$$A = \begin{Vmatrix} A_1 & A_2 \\ 0 & A_3 \end{Vmatrix},$$

where A_1 is $s \times s$, and if $J(\xi) = AM(\xi)A'$, then, with an obvious notation, we have

$$(1.7) \qquad J^*(\xi) = A_1 M^*(\xi) A_1'.$$

Suppose $\bar{G}$ is a group of linear transformations on Ω of the form (1.6) with $A_1 = I$ and $A_2 = 0$ (so that $\bar{G}$ leaves $\theta^{(1)}$ fixed), where for each $\bar{g}$ in $\bar{G}$ there is a transformation g on $\mathfrak{X}$ for which

$$(1.8) \qquad \theta' f(x) = (\bar{g}\theta)' f(gx)$$

for all x and θ. Suppose also that $G = \{g\}$ satisfies the usual conditions of the transformation group in the invariance theory in statistics (see, e.g., Kiefer [7]). Then, as proved by the author in Theorem 3.3 of [9], we have

Complete class invariance theorem. Under the above conditions, the class of designs ξ which are G-invariant, i.e., for which

$$\xi(gB) = \xi(B) \quad \text{for all } g \text{ and } B, \tag{1.9}$$

is essentially complete for linear estimation of $\theta^{(1)}$. This is a convex set of measures, invariant under G.

Fundamental in the proof of this result is the following lemma (Lemma 3.2 of [9]), which we shall also need in the next section:

LEMMA 1. *For $j = 1, 2, \cdots, r$, let C_j be $s \times (k - s)$, let D_j be positive definite symmetric $s \times s$, and suppose $\lambda_j > 0$, $\sum \lambda_j = 1$. Then*

$$[\sum \lambda_j C_j][\sum \lambda_j D_j]^{-1}[\sum \lambda_j C_j'] \leqq \sum \lambda_j C_j D_j^{-1} C_j', \tag{1.10}$$

with equality if and only if the matrix $C_j D_j^{-1}$ is the same for all j. (The extension of Lemma 1 to the case of singular D_j's is discussed shortly before the statement of Theorem 3.)

In the case where we are interested in a specific optimality criterion, the group G in the above theorem can be more general; for example, in the case of D-optimality the A_1's do not have to be the identity, but only a group of matrices of determinant one for which the usual invariance theorem in statistics holds. We obtain (Theorem 4.3 of [9])

Invariance theorem for D-optimality. Under the above conditions, there is a G-invariant ξ which is D-optimum for $\theta^{(1)}$.

This is proved by using the previous invariance theorem and also the following trivial lemma, which we shall also have occasion to use in the next section:

LEMMA 2. *If A and B are nonnegative definite symmetric $s \times s$ matrices, then* $-\log \det (\lambda A + (1 - \lambda)B)$ *is convex in λ for $0 \leqq \lambda \leqq 1$, and is strictly convex unless $A = B$ or $A + B$ is singular.*

Invariance theorems for other optimality criteria are considered in [9]. The invariance theorem for D-optimality is extremely useful in problems like those considered in Section 4 and Part II of the present paper, where it enables us to limit our search to suitably symmetrical ξ rather than to all of Ξ.

Randomization. As pointed out by the present author [8] and [9], in the exact small sample theory it can happen that, in terms of certain criteria (especially in problems of hypothesis testing), some randomized design may be better than any of the nonrandomized designs which are customarily employed in block experiments (this does not merely refer to the classical reason for employing randomization in such experiments). In the considerations of the present paper, we need not be concerned with randomized designs, since they are not needed. The reason for this is that a special case of Lemma 1 says that

$$\lambda M(\xi_1) + (1 - \lambda)M(\xi_2) \geqq [\lambda M^{-1}(\xi_1) + (1 - \lambda)M^{-1}(\xi_2)]^{-1} \tag{1.11}$$

for $0 < \lambda < 1$; if $M(\xi_1)$ and $M(\xi_2)$ are nonsingular (the singular case requiring only obvious modifications), equality can hold if and only if $M(\xi_1) = M(\xi_2)$. The right side of (1.11) is proportional to the inverse of the covariance matrix when ξ_1 is used with probability λ and ξ_2 is used with probability $(1 - \lambda)$; the left side is proportional to the information matrix of the nonrandomized design $\lambda\xi_1 + (1 - \lambda)\xi_2$; thus, the latter design is always at least as good as, and usually better than, the former (randomized) design, for linear estimation.

The number of points needed in the support of an optimum design. Elfving [4] and Chernoff [1] gave upper bounds on the number of points needed to support a design which minimizes the *average* variance of the b.l.e.'s (i.e., the trace of $M^{(1)}(\xi)$); Chernoff's result gives $s(2k - s + 1)/2$ as an upper bound in the general case. Chernoff's geometrical argument can be duplicated in the case of the generalized variance; for example, when $s = k$ this amounts to noting that the $M(\xi)$ can be considered as a closed convex set in Euclidean $k(k + 1)/2$-space with extreme points obtainable from ξ's with single points for support, and that $\det(aM)$ is increasing in $a > 0$ if M is positive definite, so that any D-optimum $M(\xi)$ must be a boundary point of the set. The bound in the case $s = 1$ is obtained by a different method in [10]. An alternative but less direct approach to obtain the bound $s(2k - s + 1)/2$ is to use Stone's characterization [13] of this as the number of points needed to support a design which maximizes $\det L_\lambda^*(\xi)$ where $L_\lambda(\xi) = I + \lambda M(\xi)$; letting λ go to infinity gives the desired result.

The bound obtained in this fashion can be very poor, as can be seen in the examples of Part II. A method of sharpening the bound slightly when $s = 1$ is given in [10]. As discussed (with slight inaccuracy) in [9] and [11], a method for sharpening the results in many cases is to count the dimension $H - 1$ (say) of the range of the set of convex linear combinations of the functions $f_i f_j$, $i \leqq j$, and then to note that any $M(\xi)$ can be achieved by a ξ whose support is H or fewer points.

Even this result gives a poor bound in many cases, as can be seen in the examples of Part II where, as is often the case, there exists a D-optimum design for θ with support on k points (obviously the minimum number possible). In the case of polynomial regression of degree $k - 1$ or less on a real interval ($f_i(x) = x^{i-1}$), it is an old and well known fact that any $M(\xi)$ (i.e., any set of values for the moments up to the $2(k - 1)$st) can be achieved by a ξ with support on at most k points, and this has been sharpened by the present author [9] to state that the minimal (essentially) complete class of admissible designs consists of those ξ whose support consists of at most k points, at most $k - 2$ of which are in the interior of the interval. A difficult and important mathematical problem is to extend such results as the well known ones just cited on the moment problem to the case of other f_i, so as to characterize for any given F the maximum number of points needed for the support of ξ's in an essentially complete class. The results for polynomials in one variable, which depend on certain properties of orthogonal polynomials, are not directly extendable.

A final problem in this area is to combine the invariance results with those just discussed. Thus, it can happen in some settings that there is an optimum design (in some sense) on P points without there being a G-invariant optimum design on P points. How many points are needed for the latter?

The nonlinear case. If the regression function is not linear in θ as in (1.2), it is still possible to obtain relevant asymptotic results. Thus, the results of the next section can be applied in exactly the way Chernoff's [1] are for the average variance, in such cases. When $s = 1$, the average and generalized variances of course coincide, and the results of Section 2 of [10] and of the present paper thus yield a computational algorithm for the problems treated by Chernoff; such an algorithm for minimizing the average variance when $s > 1$ can be found in Section 4 of [10].

The author thanks Professor J. Wolfowitz for helpful discussions.

2. The main results. Since the case where a D-optimum ξ for $\theta^{(1)}$ has *nonsingular* $M(\xi)$ is slightly easier to handle than is the singular case, and since the non-singular case yields sharper results, we shall treat this case first, although we now make some definitions which apply generally. We shall say that ξ^* yields a global minimum of det $M^{(1)}(\xi)$, or is D-optimum for $\theta^{(1)}$, if (1.3) is satisfied; of course, $M^{(1)}(\xi)$ is well defined and finite if $\theta^{(1)}$ is estimable under ξ, whether or not $M(\xi)$ is nonsingular; we define det $M^{(1)}(\xi) = \infty$ if $\theta^{(1)}$ is not estimable under ξ. Thus, in this formula as well as in those which follow, determinants and inverses can always be computed in the case where $M(\xi)$ is singular by computing them with $M(\xi)$ replaced by $M(\xi) + \lambda I$ and then letting λ approach zero. Equation (1.3) can of course be written (using (1.5)) as

$$\text{(2.1)} \quad \det M^*(\xi^*) = \det M(\xi^*)/\det M_3(\xi^*) = \max_\xi \det M^*(\xi);$$

so that (1.3) can be rephrased to state that ξ^* yields a *global maximum* of det $M^*(\xi)$. We shall say that ξ^* yields a *local maximum* of det $M^*(\xi)$ if det $M^*(\xi^*) > 0$ and

$$\text{(2.2)} \qquad \frac{\partial}{\partial \alpha} \log \det M^*([1 - \alpha]\xi^* + \alpha\xi) \,|_{\alpha=0+} \leqq 0 \quad \text{for all } \xi.$$

We generalize the definition of (1.4) in the case $s < k$ to define

$$\text{(2.3)} \qquad d(x, \xi) = f(x)'M^{-1}(\xi)f(x) - f^{(2)}(x)'M_3^{-1}(\xi)f^{(2)}(x);$$

the form of this in the case of singular $M(\xi)$ will be seen more explicitly, later (in the functions D and $\bar{D}$ introduced below; $D(x, \xi)$ is the direct analogue of $d(x, \xi)$ in the singular case, but is not the appropriate function to yield an analogue of Theorem 1, as we shall see). If $M(\xi)$ is non-singular, the integrals with respect to ξ of the two terms on the right side of (2.3) are easily seen to be k and $k - s$. Hence,

$$\text{(2.4)} \qquad \int d(x, \xi)\xi(dx) = s$$

(this is similarly true whenever $\theta^{(1)}$ is estimable, whether or not $M(\xi)$ is nonsingular, as we shall see just above (2.18)), and thus

$$\max_x d(x, \xi) \geqq s. \tag{2.5}$$

Consider now the problem of determining ξ^* so that

$$\max_x d(x, \xi^*) = \min_\xi \max_x d(x, \xi). \tag{2.6}$$

It follows from (2.5) that a *sufficient* condition for ξ^* to satisfy (2.6) is

$$\max_x d(x, \xi^*) = s. \tag{2.7}$$

(Obviously, a necessary but not sufficient condition for ξ^* to satisfy (2.7) is that ξ^* give unit measure to the set of x for which $d(x, \xi^*) = s$.)

We now prove

THEOREM 1. *If $M(\xi^*)$ is nonsingular, equations* (2.1) (*D-optimality of ξ^* for $\theta^{(1)}$*), (2.2), (2.6), *and* (2.7) *are equivalent.*

PROOF. Clearly, (2.1) implies (2.2), and we have already seen that (2.7) implies (2.6). We first show that (2.2) implies (2.7). Denoting by $M_{ij}(\xi)$ the cofactor in $M(\xi)$ of $m_{ij}(\xi)$ and by $M^{ij}(\xi)$ the (i, j)th element of $M^{-1}(\xi)$, we have, as in [11], that, for $M(\xi^*)$ nonsingular (sometimes omitting the argument ξ^* for typographical ease),

$$\begin{aligned}
&\frac{\partial}{\partial\alpha} \log \det M([1-\alpha]\xi^* + \alpha\xi) \mid_{\alpha=0+} \\
&\quad = \det M^{-1}(\xi^*) \sum_{i,j} \frac{\partial \det M}{\partial m_{ij}} \frac{\partial m_{ij}([1-\alpha]\xi^* + \alpha\xi)}{\partial\alpha}\Bigg|_{\alpha=0+} \\
&\quad = \det M^{-1}(\xi^*) \sum_{i,j} \left(\frac{\partial}{\partial m_{ij}} \textstyle\sum_q m_{iq} M_{iq}\right) [m_{ij}(\xi) - m_{ij}(\xi^*)] \\
&\quad = \det M^{-1}(\xi^*) \sum_{i,j} M_{ij}(\xi^*)[m_{ij}(\xi) - m_{ij}(\xi^*)] \\
&\quad = \sum_{i,j} m^{ij}(\xi^*) m_{ij}(\xi) - k.
\end{aligned} \tag{2.8}$$

(A somewhat neater derivation proceeds by letting $BM(\xi^*)B' = I$ and $BM(\xi)B' = D$, a diagonal matrix; the left side of (2.2) is immediately seen to be $\text{tr}D - k$, and we have $\text{tr}D = \text{tr}[BM(\xi)B'] = \text{tr}[(B'B)M(\xi)] = \text{tr}[M^{-1}(\xi^*)M(\xi)]$.) Similarly, writing $M_3^{-1} = \{\bar{m}^{ij}; i, j > s\}$, we have

$$\frac{\partial}{\partial\alpha} \log \det M_3([1-\alpha]\xi^* + \alpha\xi) \mid_{\alpha=0+} = \sum_{i,j>s} \bar{m}^{ij}(\xi^*) m_{ij}(\xi) - (k - s). \tag{2.9}$$

Hence, (2.2) can be written as

$$\begin{aligned}
&\text{tr}\,[M^{-1}(\xi^*)M(\xi)] - \text{tr}[M_3^{-1}(\xi^*)M_3(\xi)] \\
&\quad = \sum_{i,j} m^{ij}(\xi^*) m_{ij}(\xi) - \sum_{i,j>s} \bar{m}^{ij}(\xi^*) m_{ij}(\xi) \leqq s \text{ for all } \xi.
\end{aligned} \tag{2.10}$$

In particular, if ξ gives measure 1 to the point x, the left side of the inequality (2.10) becomes $d(x, \xi^*)$. Thus, (2.2) implies (2.7).

Conversely, if (2.7) holds, we have (2.10) for every ξ which gives measure one to a single point. Since $m_{ij}(\xi)$, and thus the left side of (2.10), is linear in ξ, we obtain (2.10) for all ξ. Thus, (2.7) implies (2.2).

Now, a design ξ^* satisfying (2.7) always exists (since a ξ^* satisfying (2.1) exists by compactness, and we have seen that (2.1) implies (2.7)). We conclude from (2.5) that (2.6) implies (2.7).

It remains to prove that (2.2) implies (2.1). From Lemma 1 of the previous section, we have, for $0 < \alpha < 1$,

$$M^*([1 - \alpha]\xi^* + \alpha\xi) \geqq (1 - \alpha)M^*(\xi^*) + \alpha M^*(\xi). \tag{2.11}$$

From (2.11) and Lemma 2, we obtain

$$\begin{aligned} -\log \det M^*([1 - \alpha]\xi^* + \alpha\xi) & \\ &\leqq -\log \det [(1 - \alpha)M^*(\xi^*) + \alpha M^*(\xi)] \\ &\leqq -(1 - \alpha) \log \det M^*(\xi^*) - \alpha \log \det M^*(\xi). \end{aligned} \tag{2.12}$$

(We shall later, but not now, use the fact that the first inequality is strict unless $M_2(\xi^*)M_3^{-1}(\xi^*) = M_2(\xi)M_3^{-1}(\xi)$, and that the second one is strict unless $M^*(\xi^*) = M^*(\xi)$.) Thus, log det $M^*([1 - \alpha]\xi^* + \alpha\xi)$ is concave in α, and thus this function has positive derivative at $\alpha = 0$ if det $M^*(\xi) >$ det $M^*(\xi^*)$; this last inequality holds for some ξ if (2.1) does not hold, and hence (2.2) is also violated in this case. This completes the proof of Theorem 1.

We now turn to the case where $\theta^{(1)}$ is estimable but $M(\xi^*)$ is singular. If we try to duplicate the proof of Theorem 1, we find, exactly as before, that (2.1) and (2.2) are equivalent. What happens to the rest of the proof is most clearly seen by considering a linear transformation of θ into $(A')^{-1}\theta$, where A is of the form (1.6); this transforms $M^*(\xi)$ into the form (1.7), and thus leaves unchanged the various criteria considered in Theorem 1. We shall use such a transformation in order better to display what occurs. In all that follows, ξ^* is fixed and $\theta^{(1)}$ is estimable under ξ^*. We can choose A of the form (1.6) so that $J(\xi^*) = AM(\xi^*)A'$ is a diagonal matrix with its first $s + r$ diagonal elements unity and the rest zero. For a fixed ξ, we can at the same time choose A so that

$$J(\xi) = \begin{Vmatrix} J_1 & J_2 & J_4 & 0 \\ J_2' & J_3 & J_6 & 0 \\ J_4' & J_6' & J_5 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix} \tag{2.13}$$

where J_1 is $s \times s$, J_3 is $r \times r$, J_5 is $p \times p$, J_5 is nonsingular, $J_6 = 0$, and J_1, J_3, and J_5 are diagonal.

If we try to go through the proof of Theorem 1, duplicating the arguments in the present case, we obtain

$$(2.14)\quad \begin{aligned} &J^*([1-\alpha]\xi^* + \alpha\xi) \\ &\qquad = (1-\alpha)I + \alpha D - \alpha^2 J_2([1-\alpha]I + \alpha J_3)^{-1}J_2' - \alpha J_4 J_5^{-1} J_4', \end{aligned}$$

and thus, in place of (2.10),

$$(2.15)\quad \sum_{j \leqq s} \mu_{jj}(\xi) - \rho(\xi, \xi^*) \leqq s,$$

where we denote the elements of $J(\xi)$ by $\mu_{ij}(\xi)$ and where $\rho(\xi, \xi^*)$ is the trace of $J_4 J_5^{-1} J_4'$. To put (2.15) in a form which more closely resembles (2.10), write $\mu^{ij}(\xi^*)$ for the elements of the inverse of the upper left $(r + s) \times (r + s)$ submatrix $\bar{J}(\xi^*)$ of $J(\xi^*)$, and $\bar{\mu}^{ij}(\xi^*)$, $s < i, j \leqq s + r$, for the elements of the inverse of $\{\mu_{ij}(\xi^*), s < i, j \leqq s + r\}$. Then (2.15) becomes

$$(2.16)\quad \begin{aligned} &\operatorname{tr}[\bar{J}^{-1}(\xi^*)\hat{J}(\xi)] - \operatorname{tr}[J_3^{-1}(\xi^*)\hat{J}_3(\xi)] \\ &\quad = \operatorname{tr}[\bar{J}^{-1}(\xi^*)\bar{J}(\xi)] - \operatorname{tr}[J_3^{-1}(\xi^*)J_3(\xi)] - \operatorname{tr}[J_1^{-1}(\xi^*)J_4(\xi)J_5^{-1}(\xi)J_4'(\xi)] \\ &\quad = \sum_{i,j \leqq s+r} \mu^{ij}(\xi^*)\mu_{ij}(\xi) - \sum_{\substack{s<i,j \\ \leqq s+r}} \bar{u}^{ij}(\xi^*)\mu_{ij}(\xi) - \rho(\xi, \xi^*) \leqq s, \end{aligned}$$

where $J_7' = (J_4' J_6')$ and $\hat{J}(\xi) = \bar{J}(\xi) - J_7(\xi)J_5^{-1}(\xi)J_7'(\xi)$ is proportional to the information matrix of ξ for the first $s + r$ components of $(A')^{-1}\theta$ (analogous to J^* for the first s components) and $\hat{J}_3 = J_3 - J_6 J_5^{-1} J_6'$. In fact, without requiring $J(\xi^*)$ and the $J_i(\xi)$ to be of these special forms which facilitated the computation of (2.15), we clearly have (2.16) whenever $AM(\xi^*)A'$, of rank $r + s$, has zeros outside of the upper left hand $(r + s) \times (r + s)$ matrix, where J_5 is no longer necessarily of full rank (so that we can take it to be $(k - s - r) \times (k - s - r)$) and $\rho(\xi, \xi^*)$ is again the trace of the product of $J_1^{-1}(\xi^*)$ by the matrix $\lim_{\lambda \to 0} J_4(\xi)[J_5(\xi) + \lambda I]^{-1} J_4'(\xi)$; the matrix $\hat{J}$ has the same meaning as above. Thus, for a given ξ^* and A, the same formula (2.16) holds for all ξ. In fact, it is easy to give an invariant, geometric definition of the left side of (2.16), as we shall see in Section 3, and we note here that the first form on the left side of (2.16) could have been obtained by using (2.2) and (2.10) on $\alpha\hat{J}(\xi) + (1 - \alpha)\bar{J}(\xi^*)$ with k replaced by $r + s$. If we denote by $^{\#}$ the operation $*$ of (1.5) when $k = r + s$, the essence of the matter is that, in the singular or nonsingular case,

$$(\hat{M})^{\#} = M^*.$$

This is easy to prove directly or in terms of the J's. We hereafter denote the expression on the left side of inequality (2.16) by $D(\xi, \xi^*)$; we also write $\bar{D}(\xi, \xi^*) = D(\xi, \xi^*) + \rho(\xi, \xi^*)$; this is the left side of (2.16) ignoring the $\rho(\xi, \xi^*)$ term. We also denote by $D(x, \xi^*)$ and $\bar{D}(x, \xi^*)$ and $\rho(x, \xi^*)$ the corresponding expressions when ξ gives measure one to the point x.

If $M(\xi^*)$ is singular, the functions $\bar{D}$ and ρ depend on the choice of A, but the function D does not.

It is thus clear from (2.16) that, in analogy to the implication of (2.2) by (2.7) in Theorem 1, (2.2) is now implied by the *first* or the *third* of the following four statements, and implies the *first* and *second*:

$$
\begin{aligned}
&\text{(a)}\ \max_\xi D(\xi, \xi^*) = s,\\
&\text{(b)}\ \max_x D(x, \xi^*) = s,\\
&\text{(c)}\ \max_\xi \bar{D}(\xi, \xi^*) = s,\\
&\text{(d)}\ \max_x \bar{D}(x, \xi^*) = s.
\end{aligned}
\tag{2.17}
$$

Moreover, (2.17) (c) and (d) are clearly equivalent.

By using the transformation A which makes $AM(\xi^*)A'$ the identity of order $(r + s)$ together with zeros (as above) and writing $g(x) = Af(x)$, we see that the first $r + s$ components of g are orthonormal functions with respect to ξ^*, while the other components vanish on a set of unit ξ^* measure. Hence, $D(x, \xi^*) = \sum_1^s g_i^2(x) - \rho(x, \xi^*)$, where $\rho(x, \xi^*) = 0$ on the support of ξ^*. Hence, $D(\xi^*, \xi^*) = \bar{D}(\xi^*, \xi^*) = s$, and thus, analogous to (2.5), we have

$$
\begin{aligned}
&\max_\xi \bar{D}(\xi, \xi^*) = \max_x \bar{D}(x, \xi^*)\\
&\geqq \max_\xi D(\xi, \xi^*) \geqq \max_x D(x, \xi^*) \geqq s,
\end{aligned}
\tag{2.18}
$$

for *every* ξ^* (optimum or not). Thus, if in analogy to (2.6) we set ourselves the problem of determining ξ^* so that

$$
\begin{aligned}
&\text{(a)}\ \max_\xi D(\xi, \xi^*) = \min_{\xi'}\max_\xi D(\xi, \xi'),\\
&\text{or}\\
&\text{(b)}\ \max_x D(x, \xi^*) = \min_{\xi'}\max_x D(x, \xi'),\\
&\text{or}\\
&\text{(c)}\ \max_\xi \bar{D}(\xi, \xi^*) = \min_{\xi'}\max_\xi \bar{D}(\xi, \xi'),\\
&\text{or}\\
&\text{(d)}\ \max_x \bar{D}(x, \xi^*) = \min_{\xi'}\max_x \bar{D}(x, \xi'),
\end{aligned}
\tag{2.19}
$$

it follows from (2.18) that (2.17) (a) (resp., (b), (c), (d)) is a sufficient condition for (2.19) (a) (resp., (b), (c), (d)) to be satisfied. (Of course, (2.19) (c) and (d) are equivalent.) Moreover, from (2.16) we see that there exists at least one ξ^* (namely, any satisfying (2.1)) which satisfies (2.17) (a) and (b), so that these two conditions are in fact equivalent to (2.19) (a) and (b), respectively.

We summarize our results:

THEOREM 2. *If* $\theta^{(1)}$ *is estimable under* ξ^*, *equations* (2.1), (2.2), (2.17) (a), *and* (2.19) (a) *are equivalent. Moreover,* (2.1) (*and thus any of the above*) *implies* (2.17) (b), *which is equivalent to* (2.19) (b), *while* (2.1) *is implied by* (2.17) (c) (*or, equivalently,* (d)).

An addition to this theorem, which simplifies the use of (2.17) (b), will be found in Section 3. The fact that (2.17) (c) (or equivalently, (d)) implies (2.19) (c) (or, equivalently, (d)) has not been stated as part of Theorem 2 since the latter two do not have the same intrinsic interest that (2.6) does when $s = k$. The second sentence of the theorem is not of primary interest, but (2.17) (b) is useful in *eliminating* various ξ^*'s from optimality considerations; for example, it can be used in certain problems to show that a D-optimum design cannot have such a simple structure as any of those encountered in the examples of Part II or the first example of Section 4. Of course, (2.17) (d) is a useful *sufficient* condition for D-optimality. Of primary interest is the question of whether or not (2.17) (b), (c), or (d) is equivalent to (2.17) (a), since (2.17) (b) or (d) would seem on the surface to be a more natural analogue of (2.7) than is (2.17) (a). Unfortunately (from the viewpoint of computations as well as of esthetics!), the answer in general is "no". This is easy to see by examples in the case of either of the three criteria, and we shall content ourselves here with seeing why (2.17) (b) need not entail (2.17) (a).

To this end, suppose $k = 2$, $s = 1$, and that $\mathfrak{X}$ consists of three points, with $f(x_1)' = (1,0)$, $f(x_2)' = (0,1)$, and $f(x_3)' = (b, 1)$ with $b^2 > 4$. Let ξ^* give measure 1 to x_1. Then (2.17) (b) is easily seen to be satisfied. However, if $\xi(x_2) = \xi(x_3) = \frac{1}{2}$, we have $D(\xi, \xi^*) = b^2/4 > 1$. The difficulty is really that we have lost the linearity which permitted us to go from (2.7) to (2.10), the "convexity" of Lemma 1 working in the wrong direction here.

Needless to say, there is no general equivalence of (2.17) (c) and (d) to (2.19) (c) and (d).

We end this section with a description of the set of D-optimum ξ's. It is no longer the case when $s < k$, as it was when $s = k$ (treated in [11]), that $M(\xi)$ is the same for all D-optimum ξ. From the concavity of log det $M^*([1 - \alpha]\xi^* + \alpha\xi)$ proved in (2.12) (which is valid whether or not $M(\xi)$ and $M(\xi^*)$ are nonsingular), it is clear that, if ξ^* and ξ both maximize det $M^*(\xi)$, then so does $[1 - \alpha]\xi^* + \alpha\xi$ for $0 \leqq \alpha \leqq 1$; i.e., the set of D-optimum ξ's is convex. Suppose now that $M(\xi^*)$ is nonsingular and ξ^* is D-optimum, and write $M^*(\xi^*) = R$, $M_2(\xi^*)M_3^{-1}(\xi^*) = E$, and

$$B = \begin{Vmatrix} I & -E \\ 0 & I \end{Vmatrix}. \tag{2.20}$$

If ξ is also D-optimum and $M(\xi)$ is nonsingular, we must have equality in (2.12) (otherwise the parenthetical remark following (2.12) implies that $(\xi + \xi^*)/2$ would be better). But then $M^*(\xi) = R$, $M_2(\xi)M_3^{-1}(\xi) = E$, and hence

$$BM(\xi)B' = \begin{Vmatrix} R & 0 \\ 0 & M_3(\xi) \end{Vmatrix}. \tag{2.21}$$

Conversely, if for some T we have

$$M(\xi) = B^{-1}\begin{pmatrix} R & 0 \\ 0 & T \end{pmatrix}(B^{-1})', \tag{2.22}$$

then $M^*(\xi) = R$, $M_2(\xi)M_3^{-1}(\xi) = E$, and hence ξ is D-optimum.

If $M(\xi^*)$ is singular, the characterization of (2.22) must be modified slightly. In Lemma 1 with $r = 2$, suppose the C_j and D_j are of the form

$$D_1 = \begin{Vmatrix} Q_1 & 0 & 0 \\ 0 & Q_3 & 0 \\ 0 & 0 & 0 \end{Vmatrix}, \qquad C_1 = \|L_1\, L_3\, 0\|, \qquad D_2 = \begin{Vmatrix} Q_2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_4 \end{Vmatrix}, \qquad C_2 = \|L_2\, 0\, L_4\|,$$

where the Q_i are nonsingular. Then the conclusion is modified to state that equality holds if and only if $L_1Q_1^{-1} = L_2Q_2^{-1}$. If the M_3's and M_2's are reduced to the form of the D's and C's above through a simultaneous diagonalization, we see easily that the conclusions of the previous paragraph are still valid if $M(\xi^*)$ is nonsingular. If $M(\xi^*)$ is singular, the modified conclusions can be stated in several ways, perhaps the simplest being that, if $J(\xi^*)$ is of the form prescribed above (2.13) (nothing special being assumed about the form of $J(\xi)$), then $J_2(\xi) = 0$.

Writing out the form of (2.22), we obtain

THEOREM 3. *The set of D-optimum ξ's is convex. If ξ^* is D-optimum and $M(\xi^*)$ is nonsingular, then the set of all D-optimum ξ's consists of those ξ's for which $M(\xi)$ is of the form*

$$M(\xi) = \begin{Vmatrix} R + ETE' & ET \\ TE' & T \end{Vmatrix}, \tag{2.23}$$

where $R = M^*(\xi^*)$, $E = M_2(\xi^*)M_3^{-1}(\xi^*)$, and T is arbitrary. If ξ^* is optimum and if $J(\xi^*)$ is as prescribed above (2.13), *then ξ is optimum if and only if $J_2(\xi) = 0$ and $J^*(\xi) = I$.*

In any problem where at least one optimum ξ^* exists for which $M(\xi^*)$ is nonsingular, the characterization of (2.23) can be used. The characterization of the final sentence of the theorem in the case of singular $M(\xi^*)$ can easily be given a geometric formulation in the manner of Section 3.

Of course, $M^*(\xi)$ is the same for all D-optimum ξ, which thus all perform identically for problems of linear estimation of $\theta^{(1)}$. All D-optimum (for $\theta^{(1)}$) designs are thus admissible for linear estimation of $\theta^{(1)}$, but clearly a design can be D-optimum for $\theta^{(1)}$ and inadmissible for linear estimation of the full vector θ. The designs of the form (2.23) which are also admissible for linear estimation of θ are easily characterized (see [9]) as those for which T is maximal in the sense that if a design of the form (2.23) exists with T replaced by $\bar{T}$ and with $\bar{T} \geqq T$, then $\bar{T} = T$.

3. Other forms and relationship to previous results. First suppose $M(\xi^*)$ is nonsingular and that $AM(\xi^*)A' = I$, where A is of the form (1.6). A trivial computation yields

$$M^{-1}(\xi^*) - \begin{pmatrix} 0 & 0 \\ 0 & M_3^{-1}(\xi^*) \end{pmatrix} = \begin{pmatrix} A_1' \\ A_2' \end{pmatrix} (A_1\, A_2). \tag{3.1}$$

Now, $(A_1\, A_2)\, M(\xi^*)(0\; A_4)' = 0$, and since A_4 is nonsingular this means that $(A_1 A_2)(M_2'\,(\xi^*)M_3\,(\xi^*))' = 0$. Thus, the expression (3.1) is of the form $\beta'(\beta\, M(\xi^*)\beta')^{-1}\beta$, where β is any $s \times k$ matrix of rank s whose rows are orthogonal to those of $(M_2'(\xi^*)M_3(\xi^*))$ (so that $(A_1\, A_2) = L^{-1}\beta$ where L is $s \times s$ and $LL' = \beta\, M(\xi^*)\beta'$). Writing $\beta = (B_1\, B_2)$ where B_1 is $s \times s$, we can write

$$\beta M(\xi^*)\beta' = B_1 M_1(\xi^*)B_1' - B_2 M_3(\xi^*)B_2'.$$

We can now give a geometric description of $d(x, \xi)$. Let $g_i(x) = \sum_j \beta_{ij} f_j(x)$, $1 \leqq i \leqq s$, be linearly independent with respect to (i.e. on, the support of) ξ^*, and also orthogonal (ξ^*) to all $k - s$ functions of $f^{(2)}$. Writing

$$g(x)' = (g_1(x), \cdots, g_s(x))$$

and

$$(g_i\,, g_j)_\xi = \int g_i(x) g_j(x) \xi(dx), \tag{3.2}$$

and denoting by $G(\xi)$ the matrix $\{(g_i\,, g_j)_\xi\}$, we have

$$d(x, \xi^*) = g(x)' G^{-1}(\xi^*) g(x). \tag{3.3}$$

In particular, if the g_i are chosen to be mutually orthogonal (ξ^*), we obtain

$$d(x, \xi^*) = \textstyle\sum_{i=1}^s g_i^2(x)/(g_i\,, g_i)_{\xi^*}. \tag{3.4}$$

Thus, for example, we obtain (3.4) if we let $\beta_{ii} = 1$ and choose the other β_{ij} so as to minimize (for each i)

$$\int [f_i(x) + \sum_{j \neq i} \beta_{ij} f_j(x)]^2 \xi^*(dx) \tag{3.5}$$

or, with $\beta_{ij} = 0$ for $j < i$, so as to minimize (for each i)

$$\int [f_i(x) + \sum_{j > i} \beta_{ij} f_j(x)]^2 \xi^*(dx). \tag{3.6}$$

In the case of (3.5) (resp., (3.6)), g_i is the part of f_i orthogonal (ξ^*) to the linear space spanned by the f_j with $j \neq i$ (resp., $j > i$); i.e., g_i is f_i minus the projection (ξ^*) of f_i on that linear space.

In many examples, (3.5) will be the more convenient form to use; for, if the components of $f^{(1)}$ enter the problem symmetrically, it will only be necessary to carry out the computation of the β_{ij} for a single value of i.

We shall now indicate the relationship of (3.4) with the choice (3.6) to the results of Section 4 of [10]. In the notation of the present paper, the approach of Section 4 of [10] is to consider, for $\lambda = (\lambda_1\,, \cdots, \lambda_s)$ with all $\lambda_i > 0$, the zero-sum two-person game with payoff function

$$K_\lambda(x, \beta) = \sum_{i=1}^s \lambda_i [f_i(x) + \sum_{j > i} \beta_{ij} f_j(x)]^2, \tag{3.7}$$

where $\beta = \{\beta_{ij}\,;\ 1 \leqq i \leqq s, i < j \leqq k\}$. It is shown there that if ξ_λ is a maximin strategy for this determined game, then the D-optimum ξ's are those ξ_λ's which maximize $\prod_{i=1}^{s} F_i(\xi_\lambda)$, where

$$F_i(\xi) = \min_{\{\beta_{ij}\}} \int [f_i(x) + \sum_{j>i} \beta_{ij} f_j(x)]^2 \xi(dx), \tag{3.8}$$

and that there is, to within a multiplicative constant, a unique value λ^* of λ at which the maximum is attained. The results of the present paper give us additional information. Write $\lambda_i(\xi) = 1/F_i(\xi)$ and $\lambda(\xi) = \{\lambda_i(\xi)\}$. Then, clearly, $d(x, \xi) = K_{\lambda(\xi)}(x, \beta(\xi))$ where $\beta(\xi)$ is minimal with respect to ξ for the payoff function $K_{\lambda(\xi)}$; hence,

$$\max_x d(x, \xi) = \max_x K_{\lambda(\xi)}(x, \beta(\xi)) \geqq K_{\lambda(\xi)}(\xi, \beta(\xi)) = s, \tag{3.9}$$

and a D-optimum ξ^* will be one *which is maximin when* $\lambda = \lambda(\xi^*)$ (assuming still that $M(\xi^*)$ is nonsingular for that ξ^*), and we will have equality in (3.9) for ξ equal to such a ξ^*. Hence, $\lambda^* = \lambda(\xi^*)$ for such a ξ^*, and the value of the game with $\lambda = \lambda(\xi^*) = \lambda^*$ is s. Thus, the essence of the matter is a fixed point theorem which we have proved. Hereafter calling the case where there exists a D-optimum ξ^* with nonsingular $M(\xi^*)$ the *regular case* (of $f^{(1)}$ relative to f), we have:

In the regular case there is a ξ' such that $\xi' = \xi_{\lambda(\xi')}$, and any such ξ' is D-optimum.

Of course, not all D-optimum ξ's need have $M(\xi)$ non-singular in the regular case. It will be easy to see how the above results must be modified in the singular case, but we note here that there are many examples where $\theta^{(1)}$ is estimable if and only if θ is estimable, and the above results of the regular case apply to such examples.

We remark that the considerations of Section 4 of [10] require only obvious modifications to apply to the case where (3.6) is replaced by (3.5), or where (3.4) is replaced by the general form (3.3).

To compare the method of Section 4 of [10] with the method which uses the results of Section 2 of the present paper (and which is described in the paragraph containing (1.4)), we consider the trivial problem treated in Example 4 of [10]. In the present notation, $k = 3$, $s = 2$, $\mathfrak{X} = [-1, 1]$, and $f_i(x) = x^{3-i}$. As in [10], we might begin by guessing that there is an optimum ξ of the form $\xi(-1) = \xi(1) = a$, $\xi(0) = 1\text{–}2a$. For such a ξ, we have $g_1(x) = x^2 - 2a$, $g_2(x) = x$, and

$$d(x, \xi) = \frac{(x^2 - 2a)^2}{2a(1 - a)} + \frac{x^2}{2a}. \tag{3.10}$$

This is a convex function of x^2 for $0 \leqq x^2 \leqq 1$, and thus has its maximum at $x^2 = 0$ or 1:

$$\max_x d(x, \xi) = \max\left(\frac{2a}{1 - 2a}, \frac{2 - 2a}{2a}\right). \tag{3.11}$$

This last expression equals 2 if and only if $a = 1/3$; this choice thus yields a D-optimum design. It can be seen that the computations here were exceedingly simple. A less trivial example will be found in Example 4.2. In such more complicated examples, it appears that the present method may often involve considerably less computation than that of [10].

In the case $s = 1$, we can take $\lambda = \lambda_1 = 1$ and write K for K_λ, in conformity with the notation of Section 2 of [10]. Our criterion $d(x, \xi^*) \leqq 1$ for the D-optimality of ξ^* becomes

$$\max_x K(x, \beta(\xi^*)) = K(\xi^*, \beta(\xi^*)), \tag{3.12}$$

where again $\beta(\xi^*)$ is minimal with respect to ξ^*. It follows at once from (3.12) that the game is determined and that ξ^* is maximin (and maximal with respect to $\beta(\xi^*)$) while $\beta(\xi^*)$ is minimax (and minimal with respect to ξ^*). Moreover, these assertions regarding $\beta(\xi^*)$ and ξ^* imply that $\max_x d(x, \xi^*) = 1$. Thus, the game-theoretic results of Section 2 of [10], which were proved by entirely different methods there, have been obtained at once from Theorem 1 of the present paper in the case where there exists a D-optimum ξ^* for which $M(\xi^*)$ is nonsingular. The result in the singular case can be derived with only slightly more manipulation.

In the application of our method when $s = 1$, we shall use the following notation: Suppose we are interested only in estimating θ_j, where now j need not be 1. In order to investigate the D-optimality of a ξ^* for which $M(\xi^*)$ is nonsingular, we find a column vector c of k components which is orthogonal to all columns of $M(\xi^*)$ other than the jth, and for which $c'M(\xi^*)c = 1$. Writing $\delta_j(x, \xi^*) = c'f(x)$, we then have that ξ^* is D-optimum for θ_j if and only if $\max_x |\delta_j(x, \xi^*)| = 1$.

In [11] a function space corollary of our results was stated in the case $s = k$: *There exists a ξ^* and a nonsingular $k \times k$ matrix C such that the functions $h_i(x)$, where $h(x) = Cf(x)$, are orthonormal with respect to ξ^* and* $\max_x \sum_1^k h_i^2(x) = k$. Of course, $M(\xi^*)$ is nonsingular for a D-optimum ξ^* when $s = k$, so the result there is not complicated by the possibility of singularity. One obtains a close analogue of the above result in the case $s < k$ in the regular case; in fact, given $f_1, \cdots, f_M$, the result we shall state is stronger in the case $s = N < k = M$ than is the result stated above in the case $s = k = N$. The analogue is:

COROLLARY TO THEOREM 1. *Given $\mathfrak{X}$ and $f = \{f_i, 1 \leqq i \leqq k\}$ in the regular case for $f^{(1)}$ (the first s components of f) relative to f, there exists a probability measure ξ^* on $\mathfrak{X}$ and an $s \times k$ matrix C of rank s such that the functions $h_i(x)$, where $h(x) = Cf(x)$, are orthonormal (ξ^*), are orthogonal (ξ^*) to the f_j with $j > s$, and satisfy* $\max_x \sum_1^s h_i^2(x) = s$.

We now turn to the case where $M^*(\xi^*)$ is nonsingular but $M(\xi^*)$ is singular (of course, the discussion which follows reduces to the preceding discussion if $M(\xi^*)$ is nonsingular). Let $g_i(x) = \sum_j \beta_{ij} f_j(x)$, $1 \leqq i \leqq s$, be linearly independent (ξ^*) and orthogonal (ξ^*) to all f_j, $j > s$. Let $g_i(x) = \sum_{j>s} \beta_{ij} f_j(x)$, $s < i \leqq s + r$, be a maximal set of linearly independent (ξ^*) functions of this form. Of course, each g_i is orthogonal (ξ^*) to each g_j, $1 \leqq i \leqq s < j \leqq s + r$. As in the development of (3.4), we can and do choose the g_i, $1 \leqq i \leqq s + r$, to be mutually orthogonal (ξ^*); the reader will have no difficulty in supplying the modifications needed in what follows if these functions are not so chosen.

Finally, let β_{ij}, $i > s+r$, $j > s$, be chosen so that the matrix $\{\beta_{ij}, 1 \leqq i, j \leqq k\}$ is nonsingular (where $\beta_{ij} = 0$ for $1 \leqq j \leqq s < i \leqq k$), and write $q_i(x) = \sum_{j>s} \beta_{ij} f_j(x)$, $s + r < i \leqq k$. Then the q_i are all zero on the support of ξ^*.

As in (3.4), we have

$$\bar{D}(x, \xi^*) = \textstyle\sum_{i=1}^{s} g_i^2(x)/(g_i, g_i)_{\xi^*}, \tag{3.13}$$

and we have linearity in obtaining $\bar{D}(\xi, \xi^*)$ from $\bar{D}(x, \xi^*)$:

$$\bar{D}(\xi, \xi^*) = \int \bar{D}(x, \xi^*)\xi\,(dx) = \textstyle\sum_{i=1}^{s}(g_i, g_i)_\xi/(g_i, g_i)_{\xi^*}. \tag{3.14}$$

We must still exhibit $\rho(x, \xi^*)$ and $\rho(\xi, \xi^*)$. Let $e_i(x) = \sum_{j>s+r} \gamma_{ij} q_j(x)$ be a maximal set of linearly independent (ξ) functions of this form, $1 \leqq i \leqq t$, where again for simplicity we choose the e_i to be orthogonal (ξ). Then it is easy to see that

$$\rho(\xi, \xi^*) = \sum_{\substack{j \leqq s \\ m \leqq t}} (g_j, e_m)_\xi^2/[(g_j, g_j)_{\xi^*}(e_m, e_m)_\xi]. \tag{3.15}$$

The functions $\bar{D}$ and ρ depend on the β_{ij}, but D does not. If ξ gives measure one to the point x, we obtain $\rho(x, \xi^*)$, and this takes on a particularly simple form since t must then be 0 or 1. If $t = 0$, we have $\rho(x, \xi^*) = 0$. If $t = 1$, consider the special diagonalization of $M(\xi)$ (now of rank 1) of Section 2, wherein J_1, J_3, and J_5 are diagonal and $J_6 = 0$. J_5 is a positive scalar, which we can choose to be unity. Since $J_6 = 0$, we must have $J_3 = 0$, since otherwise $J(\xi)$ would have rank >1. Similarly, at most one element, say the first, of J_1 can be other than zero, and if this element is h^2 the first element of J_4 is $\pm h$ (possibly $h = 0$), and all other elements of $J(\xi)$ except for J_5 are zero. A trivial computation yields $\bar{D}(x, \xi) = h^2$ and $D(x, \xi) = 0$ in this case (compare the example of Section 2). Thus, we see how easy it is for ξ^* to be D-optimum without (2.17)(d) being satisfied. More important, from our two results in the cases $t = 0$ and $t = 1$ we have the following sharpening of a part of the results of Section 2, which obviously shortens the computation needed to use (2.17) (b) to eliminate nonoptimum designs:

ADDITION TO THEOREM 2: *Let* $Z(\xi^*) = \{x : q_i(x) = 0, i > s + r\}$. *Then*

$$\max_x D(x, \xi^*) = \max_{x \in Z(\xi^*)} D(x, \xi^*) = \max_{x \in Z(\xi^*)} \bar{D}(x, \xi^*). \tag{3.16}$$

Equation (3.15) makes clear the lack of linearity in ξ of $\rho(\xi, \xi^*)$, which causes the complications in the singular case.

The reader will not find it difficult to write down analogues in the singular case of (3.9), (3.12), the Corollary to Theorem 1, etc.

4. Some examples.

Example 4.1. *Quadratic regression on a q-cube.* Let $\mathfrak{X}$ be the q-dimensional cube consisting of all points $x = (x_1, \cdots x_q)$ for which $-1 \leqq x_i \leqq 1$, $1 \leqq i \leqq q$. The problem of linear regression on $\mathfrak{X}$ is trivial: a D-optimum ξ is that measure which assigns measure 2^{-q} to each corner of the cube (at least for $q > 3$, there are other optimum designs,[3] since the bound H of Section 1 here becomes $H = q(q+3)/2 + 1 < 2^q$). We therefore turn to the problem of quadratic regression. The unique D-optimum ξ is well known when $q = 1$ to put equal weights on the points $x_1 = -1, 0, 1$. In what follows we restrict our attention to the case $q \geqq 2$.

It will be convenient, for the purpose of partitioning $M(\xi)$, to write the f_i in the following order: $f_1(x) = 1$; $f_{1+j}(x) = x_j^2$, $1 \leqq j \leqq q$; $f_{q+1+j}(x) = x_j$, $1 \leqq j \leqq q$; $f_i(x)$ for $2q + 2 \leqq i \leqq (q+1)(q+2)/2$ are the functions $x_p x_r$, $p < r$, in any order. Thus, $k = (q+1)(q+2)/2$, and it is easy to compute that $H = [(q+1)(q+2)(q+3)(q+4)]/24$. We shall seek an optimum ξ with support on $r = 2^{q-3}[8 + 4q + q(q-1)]$ points, of the following form: ξ assigns positive measure α to each of the 2^q corners of the cube, positive measure β to the midpoint of each of the $q2^{q-1}$ edges, and positive measure γ to the center of each of the $q(q-1)2^{q-3}$ two-dimensional (square) faces. We shall obtain such a design, and will verify its optimality, for $q = 2, 3, 4, 5$. We note that $r < H$ when $q = 2$ or 3, but that $r > H$ when $q > 3$, so that other optimum ξ exist in at least these latter cases.[4]

Although the set of points supporting the optimum ξ just described is of the same form for $q = 2, 3, 4, 5$ (the design for the case $q = 1$ is also of this form), the ratios among α, β and γ change with q. It is interesting to contrast this with the optimum ξ mentioned above for linear regression on a q-cube, or those optimum ξ's of the example of Section 6 for linear or quadratic regression on a q-simplex, where equal weights suffice in all cases. In fact, in the present example, a ξ with support of the form we are considering can no longer be optimum when $q \geqq 6$, as we shall discuss below.

For ξ of the above form, write

$$(4.1)\qquad \begin{aligned} u &= \int x_1^2 \xi\,(dx) = 2^{q-3}[8\alpha + 4(q-1)\beta + (q-1)(q-2)\gamma], \\ v &= \int x_1^2 x_2^2\, \xi\,(dx) = 2^{q-3}[8\alpha + 4(q-2)\beta + (q-2)(q-3)\gamma]. \end{aligned}$$

then

$$(4.2)\qquad M(\xi) = \begin{Vmatrix} 1 & F & 0 & 0 \\ F' & G & 0 & 0 \\ 0 & 0 & uI_q & 0 \\ 0 & 0 & 0 & vI_{q(q-1)/2} \end{Vmatrix},$$

where I_q is the $q \times q$ identity, F is a row-vector of q u's, G is a $q \times q$ matrix with

---

[3] In fact, for $q \geqq 3$, an optimum design assigning measure $1/h$ to each of h points of a proper subset of the 2^q corners can be obtained from an Hadamard matrix or orthogonal array of strength 2 which describes a design for the corresponding factorial problem with q factors at 2 levels. Here h can be taken to be $\leqq 2q$ (an easily improvable bound), so that we see again how poor the bound H can be. These results on linear regression are much simpler than the corresponding results on quadratic regression which are mentioned in footnote 5.

[4] See footnote 5 in this connection.

diagonal elements u and off-diagonal elements v, and the symbol 0 denotes any matrix of zeros. From this we obtain easily

$$(4.3)\qquad M(\xi)^{-1} = \begin{Vmatrix} a & B & 0 & 0 \\ B' & C & 0 & 0 \\ 0 & 0 & u^{-1}I_q & 0 \\ 0 & 0 & 0 & v^{-1}I_{q(q-1)/2} \end{Vmatrix}$$

where $a = [(q-1)v + u]/[(q-1)v + u - qu^2]$, each of the q elements of B is $b = -u/[(q-1)v + u - qu^2]$, and C has diagonal elements

$$c = [(q-2)v + u - (q-1)u^2]/(u-v)[(q-1)v + u - qu^2]$$

and off-diagonal elements $d = [u^2 - v]/(u-v)[(q-1)v + u - qu^2]$. Also, from (4.2), we have

$$(4.4)\qquad \det M(\xi) = u^q v^{q(q-1)/2}(u-v)^{q-1}[u + (q-1)v - qu^2].$$

Since the problem at hand is illustrative of many similar examples, we now indicate two methods for "guessing" values α, β, γ for which one can verify that $\max_x d(x, \xi) = (q+1)(q+2)/2$. Firstly, as mentioned in the introduction, we can try to maximize $\det M(\xi)$ among ξ of this form, by solving the equations $\partial \log \det M(\xi)/\partial u = \partial \log \det M(\xi)/\partial v = 0$ in the region where α, β, and γ are all positive. Secondly, we can used (4.3) to write out $d(x, \xi)$, say $d(x, \xi) = P + Q\sum_i x_i^2 + R\sum_i x_i^4 + S\sum_{i\neq j} x_i^2 x_j^2$, where P, Q, R, S are functions of u and v, and then try to determine u and v so as to make $d(x, \xi)$ have some simple form for which it is obvious that $\max_x d(x, \xi) = (q+1)(q+2)/2$; for example, we can try to find u and v such that $P = (q+1)(q+2)/2$, $R = -Q \geqq 0$, $S = 0$. Either of these approaches leads to the same formal solution in the present cases, neglecting for the moment the question of positivity of α, β, γ:

$$(4.5)\qquad \begin{aligned} u &= \frac{(q+3)}{4(q+1)(q+2)^2}\{(2q^2 + 3q + 7) + (q-1)[4q^2 + 12q + 17]^{\frac{1}{2}}\}, \\ v &= \frac{(q+3)}{8(q+2)^3(q+1)} \\ &\quad \cdot \{(4q^3 + 8q^2 + 11q - 5) + (2q^2 + q + 3)[4q^2 + 12q + 17]^{\frac{1}{2}}\}. \end{aligned}$$

For this choice of u and v we obtain after some reduction

$$(4.6)\qquad d(x, \xi) = (q+1)(q+2)/2 - c\sum_i (x_i^2 - x_i^4),$$

whose maximum over $\mathfrak{X}$ is clearly the desired value $(q+1)(q+2)/2 = k$, since c, defined just below (4.3), is easily seen to be positive. The corresponding values of α, β, and γ which are obtained from the equations (4.1) and the equation $2^{q-3}[8\alpha + 4q\beta + q(q-1)\gamma] = 1$, are

$$\alpha = 2^{-q-1}[(q-1)(q-2) - 2q(q-2)u + q(q-1)v],$$

$$\beta = 2^{-q+1}[(2q - 3)u - (q - 1)v - (q - 2)],$$
$$\gamma = 2^{2-q}[1 + v - 2u];$$

more explicitly,

$$\begin{aligned}
\alpha = {} & [2^{q+4}(q + 2)^3(q + 1)]^{-1}\{(4q^6 + 12q^5 \\
& - 25q^4 - 107q^3 + 85q^2 + 479q + 128) \\
& - (2q^2 - q - 19)q\,(q - 1)(q + 3)[4q^2 + 12q + 17]^{\frac{1}{2}}\},
\end{aligned}$$

$$\begin{aligned}
(4.7)\quad \beta = {} & [2^{q+2}(q + 2)^3(q + 1)]^{-1}\{-(4q^5 + 16q^4 - 11q^3 - 143q^2 \\
& -149q + 139) + (q + 3)(q - 1)(2q^2 + q - 15)[4q^2 + 12q + 17]^{\frac{1}{2}}\}
\end{aligned}$$

$$\begin{aligned}
\gamma = {} & [2^{q+1}(q + 2)^3(q + 1)]^{-1}\{(4q^4 + 24q^3 + 43q^2 - 24q - 119 \\
& -(q + 3)(2q^2 + 3q - 11)[4q^2 + 12q + 17]^{\frac{1}{2}}\}.
\end{aligned}$$

Thus, (4.7) provides an optimum ξ, provided that the α, β, γ given here are all nonnegative. This is the case for $q \leqq 5$, and the following is a table of numerical values:

q	α	β	γ
1	.250	.500	.000
2	.1458	.08015	.0962
3	.071975	.01895	.03280
4	.03705	.0038375	.01185
5	.01928	.0003125	.004475

For comparison, we note that, when $q = 2$, the ξ which assigns measure $\frac{1}{9}$ to each of the nine points supporting the optimum ξ, yields a value of det $M(\xi)$ which is about 15 per cent lower and a value of $\max_x d(x, \xi)$ which is about 21 per cent higher, than does the optimum design. For larger q, the comparison is even more striking.

To see what happens to the above solution when $q > 5$, it will suffice to consider the case $q = 6$. Equation (4.7) no longer gives a solution, since $\beta < 0$ (i.e., the solution can no longer be obtained by solving $\partial \log \det M(\xi)/\partial u = \partial \log \det M(\xi)/\partial v = 0$). This suggests that we look for a D-optimum ξ of the form we have been considering, but with $\beta = 0$. If, in fact, we investigate the behavior of the expression (4.4) on the region $\{\alpha \geqq 0, \beta \geqq 0\ \gamma \geqq 0\} = \{u \leqq (v + 1)/2, u \leqq (10 + 15v)/24, u \geqq (4 + 5v)/9\}$, we find that the maximum is attained at $(u, v) = ([5v' + 4]/9, v')$, where $v = v'$ is the solution between .7 and .8 of the equation $350v^3 - 190v^2 - 139v + 60 = 0$ (this last equation is obtained by solving $\partial \log \det M(\xi)/\partial v = 0$ on the line $9u = 4 + 5v$, and it is not hard to prove that this gives the desired solution). For the corresponding ξ (for which $\beta = 0$) we obtain, at $x = 0, d(0, \xi) = 3(25v' + 2)/5(1 - v')(5v' - 2) > 28 = k$. Hence, we have proved that *the best ξ of the form we have considered* (i.e., *over all choices of* α, β, γ) *is not D-optimum when* $q = 6$. The corresponding result also holds when $m > 6$, and a D-optimum design for the case $q \geqq 6$ is still unknown.[5]

Example 4.2. *The case of polynomial regression on a real interval when* $1 < s < k$. The problem of polynomial regression on a real interval was solved by Guest [5] and Hoel [6] in the case $s = k$ and by Kiefer and Wolfowitz [10, Section 3] in the case $s = 1$. The other cases are more difficult to handle. A trivial example (quadratic regression, $k = 3$, $s = 2$) was treated in the previous section and in [10], and we now illustrate the more complicated problems which can arise by considering two computationally more difficult examples for the case $s = 2 < k$. In both examples it is obvious from the outset that we are in the regular case of Section 3.

First consider the problem of estimating the quadratic and cubic regression coefficients in the case of cubic regression; i.e. $s = 2$, $k = 4$, $\mathfrak{X} = [-1, 1]$, and $f_i(x) = x^{4-i}$, $i = 1, 2, 3, 4$; we want a D optimum design for estimating θ_1 and θ_2 (the coefficients of x^3 and x^2), and the comments of Section 1 suggest that we seek one of the form $\xi(a) = \xi(-a) = \alpha/2$, $\xi(1) = \xi(-1) = (1 - \alpha)/2$, where $0 < a < 1$. We easily compute that the $g_i(x)$ of Section 3 can be taken to be $x^3 - cx$ and $x^2 - b$, where $c = (1 - \alpha + \alpha a^4)/(1 - \alpha + \alpha a^2)$ and $b = 1 - \alpha + \alpha a^2$. Writing $x^2 = u$ and $a^2 = A$, we obtain

$$(4.8) \quad d(x, \xi) = \frac{(u - b)^2}{\alpha(1 - \alpha)(1 - A)^2} + \frac{(1 - \alpha + \alpha A)}{(1 - \alpha)\alpha A(1 - A)^2} u(u - c)^2.$$

If ξ is D-optimum, we must have $d(1, \xi) = 2$; i.e.,

$$(4.9) \quad 2z^2 + (A - 1)z - 2A = 0,$$

where we have written $z = (1 - \alpha)/\alpha$. If $d(1, \xi) = 2$, we must also have $d(a, \xi) = 2$, and since the expression (4.8) is a cubic in u we will clearly have $d(x, \xi) \leqq 2$ for all x if

$$(4.10) \quad \begin{aligned} \partial d(x, \xi)/\partial u \,|_{u=A} &= 0, \\ \partial^2 d(x, \xi)/\partial u^2 \,|_{u=A} &< 0. \end{aligned}$$

The first half of (4.10) yields $z = 3A^2/(1 - 4A)$; substituting this into (4.9), we obtain, finally,

$$(4.11) \quad \begin{aligned} a &= A^{\frac{1}{2}} = [(11 - 73^{\frac{1}{2}})/12]^{\frac{1}{2}}, \\ \alpha &= (z + 1)^{-1} = (73^{\frac{1}{2}} - 5)/6, \end{aligned}$$

[5] Recently Dr. R. H. Farrell and the author have obtained optimum designs for all values of q. For $q > 5$, the support of such designs must contain points of the 3^q array which are midpoints of faces of dimension >2. The invariant designs of this form (which are not unique for $q > 2$) can always be obtained by choosing weights analogous to α, β, and γ above in such a way as to make the moments defined by the lefthand equations of (4.1) equal to the quantities defined by (4.5). The designs obtained in this way will be supported by more than H points if $q > 5$. Designs on fewer than H points (in fact, on $O(q^3)$ points) can be obtained by combining certain orthogonal arrays of strength 4. Results of the type discussed here and in footnote 3 will appear elsewhere.

and it is easy to check that (4.9) and (4.10) are satisfied by these values. Thus, (4.11) gives a D-optimum design.

Next, suppose with the same cubic setup that we only have $k = 3$ (i.e., the constant term is missing). Surprisingly, the arithmetic is now more complicated. One obtains $z^3 - (1 - 2A^2)z^2 - (2A - A^3)z - A^3 = 0$ in place of (4.9) and $(3A - 1)z^2 + (5A^3 - A^2)z + (4A^4 - 2A^3) = 0$ for the first half of (4.10), and more effort is required to solve these than in the previous case. We obtain, finally, that a D-optimum design ξ of the same structure as above is now given by

$$a = [(5 \cdot 33^{\frac{1}{2}} - 21)/24]^{\frac{1}{2}}, \quad \alpha = (3 + 33^{\frac{1}{2}})/20. \tag{4.12}$$

PART II. SIMPLEX EXPERIMENTS

5. Preliminaries. Scheffé [12] has given an interesting account of experiments in which $\mathfrak{X}$ is the *q-simplex* S_q consisting of all $(q + 1)$-vectors $(x_1, x_2, \cdots, x_{q+1})$ for which all x_i are nonnegative and $\sum_i x_i = 1$. (Scheffé uses $q - 1$ to denote the dimensionality of the simplex, but we shall find the present notation more convenient, and will adhere to it throughout.) The reader is referred to the fundamental paper [12] for discussions of the construction and use of such experiments, including modifications in the case where $\mathfrak{X}$ is only a part of S_q. We shall be concerned here with optimum properties which are possessed (or not possessed) by certain of Scheffé's designs, namely, those designs in which ξ gives measure 1 to the *(q, m)-lattice* $S_{q,m}$ consisting of those $\binom{m+q}{m}$ points of S_q all of whose coordinates are integral multiples of $1/m$, and, in particular, the design $\xi_{q,m}$ whch assigns equal measure to each of these points.

In the footnote on page 353 of [12], Scheffé mentions the desirability of investigating the optimality of his designs (in the case $s = k$ of our Section 1) in precisely the sense discussed in Part I of the present paper, i.e., in the sense of minimizing $\max_x d(x, \xi)$. We shall investigate the optimality of $\xi_{q,m}$ or certain simple modifications of it in various cases of polynomial regression on $\mathfrak{X}$. Thus, in the case where all polynomials on S_q of degree m or less are possible regression functions, the set $\{f_i\}$ of Section 1 can be chosen in various ways (see [12]) as a set of $\binom{m+q}{m}$ linearly independent polynomials of degree $\leq m$. We shall also discuss certain other cases considered by Scheffé, in which only a proper subset of the polynomials of degree m are possible.

Before proceeding to these investigations, it is necessary to verify a conjecture of Scheffé regarding designs on $S_{q,m}$:

Orthogonal polynomials and identifiability for designs on a (q, m) lattice. Scheffé makes a conjecture on page 346 of [12] which is equivalent to the statement that, for mth degree regression, any design which gives positive measure to all $\binom{m+q}{m}$

points of the (q, m) lattice $S_{q,m}$ enables all $\binom{m+q}{m}$ regression coefficients to be estimated. (He verifies this for $m = 1, 2, 3$.) We now verify this conjecture by proving the existence of a system of $\binom{m+q}{m}$ polynomials of degree $\leqq m$ such that, for any point of the lattice, there is a polynomial in the system which is not zero at that point, but which vanishes at all other points of the lattice. (This system is thus orthogonal for any design whose support is the (q, m) lattice.) Since there are exactly as many points of the lattice as there are regression coefficients, this will imply the validity of Scheffé's conjecture.[6]

Fix q. Such a system of polynomials obviously exists when $m = 1$. Suppose such a system exists when $m = M - 1$, where $M > 1$. Let p be a point of $S_{q,M}$. Since $M > 1$, there is a bounding hyperplane L of the simplex S_q on which $S_{q,M}$ is the lattice, such that $p \not\in L$. Since $T = S_{q,M} - L$ is essentially a $(q, M - 1)$ lattice, there is a polynomial Φ of degree at most $M - 1$ which vanishes everywhere on T except at p. But then, if f is a linear function which vanishes on L but not on T, the function Φf is a polynomial of degree at most M which vanishes everywhere on $S_{q,M}$ except at p. This completes the proof.

6. Quadratic regression on the q-simplex.

A D-optimum design for all coefficients in quadratic regression on the q-simplex. We shall now show that, when $\mathfrak{X}$ is the q-simplex S_q and F consists of all polynomials of degree $\leqq 2$, the design $\xi_{2,q}$ which assigns measure $2/(q + 1)(q + 2)$ to each of the points of the $(q, 2)$ lattice $S_{q,2}$ on $\mathfrak{X}$, is D-optimum. To this end, we compute $d(x, \xi_{2,q})$. This can be done directly by computing $M(\xi_{2,q})^{-1}$ and thus $f(x)'M(\xi_{2,q})^{-1}f(x)$ for the usual choice $\{f_i(x)\} = \{x_r, 1 \leqq r \leqq q + 1$ and $x_r s_s, 1 \leqq r < s \leqq q + 1\}$, but a somewhat quicker method is to note that a system of $(q + 1)(q + 2)/2$ quadratic orthonormal polynomials with respect to $\xi_{2,q}$, each of which vanishes except at one point of the lattice, consists of the functions $[2(q + 1)(q + 2)]^{\frac{1}{2}} x_i(x_i - \frac{1}{2})$, $1 \leqq i \leqq q + 1$, and the functions $[8(q + 1)(q + 2)]^{\frac{1}{2}} x_i x_j$, $1 \leqq i < j \leqq q + 1$. Hence, $d(x, \xi_{2,q})$ is just the sum of squares of these functions (see Section 3), and we obtain, denoting by $\sum'$ the summation over all j not equal to i (for fixed i),

$$\frac{2}{(q + 1)(q + 2)} d(x, \xi_{2,q}) = 4 \sum_i x_i^2(x_i - 1/2)^2 + 16 \sum_{i<j} x_i^2 x_j^2$$
$$= \sum_i (1 - \textstyle\sum' x_j)(4x_i^3 - 4x_i^2 + x_i) + 8 \displaystyle\sum_{i \neq j} x_i^2 x_j^2$$

---

[6] It will be seen that it is unnecessary to exhibit these polynomials explicitly in carrying out the inductive proof which follows, although that induction can be used to obtain them explicitly. Professor Scheffé has informed the author that I rofessor L. J. Savage had independendently constructed and communicated to him the formula for a polynomial of degree m on S_q which vanishes at all points of the (q, m) lattice except for the point $(z_1, z_2, \cdots, z_{q+1})$, where it is unity. Savage's expression is

$$\prod_{i=1}^{q+1} \left\{ [(mz_i)!]^{-1} \prod_{j=0}^{mz_i - 1} (mx_i - j) \right\}.$$

$$= 1 - \sum_i 4x_i^2(1 - x_i) - \sum_i (2x_i - 1)^2 x_i {\sum}' x_j + 8\sum_{i \neq j} x_i^2 x_j^2$$
$$= 1 - \sum_{i \neq j} x_i x_j \{4x_i + (2x_i - 1)^2\} + 8\sum_{i \neq j} x_i^2 x_j^2$$
$$= 1 - \sum_{i \neq j} x_i x_j \{2x_i + 2x_j + (2x_i - 1)^2/2 + (2x_j - 1)^2/2\} + 8\sum_{i \neq j} x_i^2 x_j^2$$
$$= 1 - \sum_{i \neq j} x_i x_j \{2(x_i - x_j)^2 + (1 - 4x_i x_j)\}.$$

The last expression in braces is always nonnegative. Hence, $d(x, \xi_{2,q}) \leqq (q + 1)(q + 2)/2$ for all x, and $\xi_{2,q}$ is indeed optimum.

It is striking to note how much simpler the treatment of the present example is, than is that of quadratic regression on the q-cube in Section 4. Unfortunately, the cases where $m \geqq 3$ are not so simple.

An optimum design for estimating only the coefficients of the quadratic terms of a quadratic on S_2. This example will illustrate the use of our theory when $1 < s < k$, and contains a good example of the type of geometric argument which is often useful. Write the f_i's in order as x_2x_3, x_1x_3, x_1x_2, x_1, x_2, x_3. We seek a design which minimizes the generalized variance of the three b.l.e.'s of coefficients of f_1, f_2, f_3. It is to be noted that any D-optimum design for this problem is also D-optimum for the problem where f_1, f_2, f_3 are replaced by x_1^2, x_2^2, x_3^2, since the transformation which takes one problem into the other is of the form (1.6).

We shall search for an optimum design among those designs $\xi^{(\alpha)}$ which, for some α, assign measure $\alpha/3$ to each vertex of S_2 and measure $(1 - \alpha)/3$ to the midpoint of each edge of S_2. Denoting by $[a, b]$ a 3×3 matrix with diagonal elements a and off-diagonal elements b, we obtain for such a design $\xi^{(\alpha)}$,

$$3M(\xi^{(\alpha)}) = \begin{Vmatrix} [\alpha/16, 0] & [0, \alpha/8] \\ [0, \alpha/8] & [1 - \alpha/2, \alpha/4] \end{Vmatrix}$$

and thus

$$\frac{1}{3} M^{-1}(\xi^{(\alpha)}) - \frac{1}{3} M_2^{-1}(\xi^{(\alpha)}) = \begin{Vmatrix} \left[\frac{8(2-\alpha)}{\alpha(1-\alpha)}, \frac{4}{1-\alpha}\right] & \left[0, \frac{-2}{1-\alpha}\right] \\ \left[0, \frac{-2}{1-\alpha}\right] & \left[\frac{2\alpha - \alpha^2}{(4-3\alpha)(1-\alpha)}, \frac{\alpha}{4-3\alpha}\right] \end{Vmatrix}$$

and (using the fact that $\sum x_i = 1$)

$$\frac{1}{3} d(x, \xi^{(\alpha)}) = \frac{(2-\alpha)\alpha}{(4-3\alpha)(1-\alpha)} \sum_i x_i^2 + \frac{2\alpha}{4-3\alpha} \sum_{i<j} x_i x_j$$
$$- \frac{4}{1-\alpha} \sum_{i \neq j} x_i^2 x_j + \frac{8(2-\alpha)}{\alpha(1-\alpha)} \sum_{i<j} x_i^2 x_j^2 + \frac{8}{1-\alpha} x_1 x_2 x_3.$$

Of course, a necessary condition for optimality is that $d(x, \xi^{(\alpha)}) = 3$ on a set of unit $\xi^{(\alpha)}$-measure. It is only necessary to check this condition at the point

(1, 0, 0), since it then follows for other relevant points from symmetry and the fact that the integral of $d(x, \xi^{(\alpha)})$ with respect to $\xi^{(\alpha)}$ is automatically 3. We obtain $\alpha = \bar{\alpha}$, where

$$\bar{\alpha} = \frac{9 - 17^{\frac{1}{2}}}{8} = .6530.$$

In order to prove that $\xi^{(\bar{\alpha})}$ is optimum, we must show that $d(x, \xi^{(\bar{\alpha})}) \leqq 3$ on S_2. First we note that if we consider the function $d(x, \xi^{(\bar{\alpha})})$ not merely on S_2 but on the whole plane $P = \{\sum x_i = 1\}$, it is obviously a quartic which is non-negative (see (3.4)) and which, on the line $x_3 = 0$, is symmetric about $(\frac{1}{2}, \frac{1}{2}, 0)$ and equal to 3 on this line at $x_1 = 0, \frac{1}{2}$, and 1. We conclude without any computation that $d(x, \xi^{(\bar{\alpha})}) \leqq 3$ on that part of the line $x_3 = 0$ which is part of S_2, and thus on the whole boundary of S_2.

Next, we compute easily that $d(x', \xi^{(\bar{\alpha})}) < 3$, where $x' = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Furthermore, it is not hard to compute that x' is a local strict maximum of $d(x, \xi^{(\alpha)})$. From this and the fact that d is positive quartic on the plane P which is $\leqq 3$ on the boundary of S_2, we conclude easily that $d(x, \xi^{(\bar{\alpha})}) \leqq 3$ on that part of any line of P through x' which is contained in S_2. Hence, $d(x, \xi^{(\bar{\alpha})}) \leqq 3$ throughout S_2, and thus $\xi^{(\bar{\alpha})}$ is D-optimum.

7. Cubic and higher regression on the q-simplex. The cases where $m \geqq 3$ are computationally much more difficult. We already know from the results of Guest [5] and Hoel [6] that, even in the case $q = 1$, any design on the (q, m) lattice (regardless of whether or not ξ assigns equal measure to the points) is *not* D-optimum when $m \geqq 3$.

For the sake of brevity, we will limit our discussion to the case $q = 2$, $m = 3$. We shall briefly discuss three different models. In the general cubic case (in Scheffé's terminology) we can take the f_i to be the ten functions x_i, $x_i x_j$, $x_i x_j(x_i - x_j)$, and $x_1 x_2 x_3$ (here $1 \leqq i < j \leqq 3$). Scheffé's "special cubic" omits the functions $x_i x_j(x_i - x_j)$. In the "cubic without 3-way effect" we shall consider the nine functions other than $x_1 x_2 x_3$; it is clear in what sense the meaning of this name is to be taken, and the physical significance of each of the three models is clear (see also [12]).

In the case of the cubic without 3-way effect, for $0 < b < \frac{1}{2}$ consider the design ξ_b which puts measure $\frac{1}{9}$ on each of the three points $x_i = 1, x_j = x_k = 0$ and each of the six points $x_i = 1 - x_j = b, x_k = 0$. It is not too difficult to compute that

$$\det M(\xi_b) = \text{const } V^{12}(1 - 4V)^3,$$

where $V = b(1 - b)$. Hence, $V = \frac{1}{5}$, or $b = (1 - 5^{-\frac{1}{2}})/2$, gives the optimum design *among designs of this structure*. It is interesting to note that *this value of b also gives the D-optimum design in the case* $q = 1$, $m = 3$, with equal weights at each of the points $x_1 = 0, b, 1 - b$, and 1.

For the general cubic, if we consider designs which assign measure $\frac{1}{10}$ to each of the nine points supporting ξ_b in the previous paragraph, and also to the point

$x_1 = x_2 = x_3 = \frac{1}{3}$, the best choice of b changes to $(1 - 3^{-\frac{1}{2}})/2$. In fact, it is far from clear that we should expect the D-optimum ξ to be of this form or to be supported by only 10 points; the situation appears to be more complex than that of quadratic regression on a square (discussed in Section 4).

A D-optimum design for the special cubic on S_2. We turn now to the case of the special cubic, where we shall show that Scheffé's design $\bar{\xi}$ which assigns measure $\frac{1}{7}$ to each of the six points of the (2, 2) lattice $S_{2,2}$ and also to the point $x_1 = x_2 = x_3 = \frac{1}{3}$, is indeed D-optimum. We cannot, in imitation of our development in Section 6, take $d(x, \bar{\xi})$ to be the sum of squares of the seven orthonormal cubic functions each of which vanishes on all but one of these seven points; for these cubics will not all be linear combinations of *only* the seven functions we began with. Rather than to compute appropriate orthogonal functions, we shall in this example compute M^{-1} directly. Writing the seven functions in the order $x_1, x_2, x_3, x_2x_3, x_3x_1, x_1x_2, x_1x_2x_3$, and denoting by $[a, b]$ a 3×3 matrix with diagonal elements a and off-diagonal elements b, and by $[c]$ a 3×1 matrix of elements c, we obtain

$$7M(\bar{\xi}) = \left\| \begin{matrix} \left[\frac{29}{18}, \frac{13}{36}\right] & \left[\frac{1}{27}, \frac{35}{216}\right] & \left[\frac{1}{81}\right] \\ \left[\frac{1}{27}, \frac{35}{216}\right] & \left[\frac{97}{1296}, \frac{1}{81}\right] & \left[\frac{1}{243}\right] \\ \left[\frac{1}{81}\right]' & \left[\frac{1}{243}\right]' & \left[\frac{1}{729}\right] \end{matrix} \right\|$$

and thus

$$\frac{1}{7} M^{-1}(\bar{\xi}) = \left\| \begin{matrix} [1, 0] & [0, -2] & [3] \\ [0, -2] & [24, 4] & [-60] \\ [3]' & [-60]' & 1188 \end{matrix} \right\|.$$

Hence, we obtain

$$\frac{1}{7}d(x, \bar{\xi}) = \sum_i x_i^2 - 4\sum_{i \neq j} x_i^2 x_j + 24 \sum_{i<j} x_i^2 x_j^2 + 8 \sum_i x_i x_1 x_2 x_3 + 6x_1x_2x_3 \tag{7.1}$$
$$- 120 x_1x_2x_3 \sum_{i<j} x_i x_j + 1188 x_1^2 x_2^2 x_3^2 .$$

The fourth term on the right is of course just $8x_1x_2x_3$. The first term on the right can be written as

$$\sum_i x_i^2 = 1 - 2\sum_{i<j} x_i x_j = 1 - 2 \sum_{\substack{i<j \\ i \neq k \neq j}} x_i x_j (x_i + x_j + x_k) \tag{7.2}$$
$$= 1 - 2 \sum_{i \neq j} x_i^2 x_j - 6 x_1 x_2 x_3 .$$

We substitute this last expression in (7.1) and, in the resulting form, substitute for the expression $-6\sum_{i \neq j} x_i^2 x_j$ the last of the following expressions:

$$-6\sum_{i\neq j} x_i^2 x_j = -6\sum_{i\neq j} x_i^3 x_j - 6\sum_{i\neq j} x_i^2 x_j(1 - x_i)$$
$$(7.3) \qquad = -6\sum_{i\neq j} x_i^3 x_j - 6 \sum_{\substack{i\neq j\\ i\neq k\neq j}} x_i^2 x_j(x_j + x_k)$$
$$= -6\sum_{i\neq j} x_i^3 x_j - 12\sum_{i<j} x_i^2 x_j^2 - 12 x_1 x_2 x_3 .$$

We obtain, finally,

$$\tfrac{1}{7}d(x,\bar{\xi}) = 1 - 6\sum_{i\neq j} x_i^3 x_j + 12\sum_{i<j} x_i^2 x_j^2 - 4x_1x_2x_3$$
$$-120 x_1 x_2 x_3 \sum_{i<j} x_i x_j + 1188 x_1^2 x_2^2 x_3^2$$
$$(7.4) \qquad = 1 - \{6\sum_{i<j} x_i x_j (x_i - x_j)^2 + 4x_1x_2x_3(1 - 27x_1x_2x_3)$$
$$+ 120 x_1 x_2 x_3 (\sum_{i<j} x_i x_j - 9 x_1 x_2 x_3)\}.$$

Each of the three terms inside the curly braces is easily seen to be nonnegative on the simplex. Hence, $d(x, \bar{\xi}) \leqq 7$ for all x in the simplex, and thus $\bar{\xi}$ is indeed D-optimum.

An optimum design for estimating only the coefficient of the cubic term of a special cubic on S_2. Scheffé showed that, among the designs which assign measure one to the set of seven points which supports the $\bar{\xi}$ of the previous example, the one which minimizes the variance of the b.l.e. of the coefficient of $x_1x_2x_3$ is the measure ξ' which assigns measure $\frac{1}{24}$ to each vertex of S_2, $\frac{4}{24}$ to the midpoint of each side of S_2, and $\frac{9}{24}$ to the centroid of S_2. We now show that, in fact, ξ' is optimum among *all* designs.

The proof is quite simple. Using the notation of the previous example, we obtain

$$(7.5) \qquad 24M(\xi') = \begin{Vmatrix} [4, 2] & [1/3, 5/6] & [1/9] \\ [1/3, 5/6] & [13/36, 1/9] & [1/27] \\ [1/9]' & [1/27]' & 1/81 \end{Vmatrix}$$

A column vector c which is orthogonal to the first six columns of $M(\xi')$ and for which $c'M(\xi')c = 1$ is given by $c' = (1, 1, 1, -8, -8, -8, 72)$. Thus, in the notation of Section 3,

$$\delta_7(x, \xi') = \sum_i x_i - 8\sum_{i<j} x_i x_j + 72 x_1 x_2 x_3$$
$$(7.6) \qquad = 1 - 8\{x_1x_2 + x_1x_3 + x_2x_3 - 9\, x_1x_2x_3\}.$$

The term in braces is easily seen to have a maximum of $\frac{1}{4}$ and a minimum of 0 on S_2. Hence, $\max_x |\delta_7(x, \xi')| = 1$, and thus (see Section 3) ξ' is optimum.

REFERENCES

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

[2] S. EHRENFELD, "On the efficiency of experimental designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 247–255.
[3] S. EHRENFELD, "Complete class theorems in experimental design," *Proceedings Third Berkeley Symposium on Mathematical Statistics and Probability*, University California Press, Berkeley 1955.
[4] G. ELFVING, "Optimum allocation in linear regression theory," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 255–262.
[5] P. G. GUEST, "The spacing of observations in polynomial regression," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 294–299.
[6] P. G. HOEL, "Efficiency problems in polynomial estimation," *Ann. Math. Stat.*, Vol. 19 (1958), pp. 1134–46.
[7] J. KIEFER, "Invariance, minimax sequential estimation, and continuous time processes," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 573–601.
[8] J. KIEFER, "On the nonrandomized optimality and randomized nonoptimality of symmetrical designs," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 675–699.
[9] J. KIEFER, "Optimum experimental designs," *J.R.S.S.* (Ser. B), Vol. 21 (1959), pp. 272–319.
[10] J. KIEFER AND J. WOLFOWITZ, "Optimum designs in regression problems," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 271–294.
[11] J. KIEFER AND J. WOLFOWITZ, "The equivalence of two extremum problems," *Can. Jnl. Math.*, Vol. 12 (1960), pp. 363–366.
[12] H. SCHEFFÉ, "Experiments with mixtures," *J.R.S.S.* (Ser. B), Vol. 20 (1958), pp. 344–360.
[13] M. STONE, "Application of a measure of information to the design and comparison of regression experiments," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 55–70.