The Annals of Mathematical Statistics
1969, Vol. 40, No. 5, 1570–1602

OPTIMAL AND EFFICIENT DESIGNS OF EXPERIMENTS[1]

BY CORWIN L. ATWOOD

University of Minnesota, Minneapolis

0. Summary. This paper consists of new results continuing the series of papers on optimal design theory by Kiefer (1959), (1960), (1961), Kiefer and Wolfowitz (1959), (1960), Farrell, Kiefer and Walbran (1965) and Karlin and Studden (1966a). After disposing of the necessary preliminaries in Section 1, we show in Section 2 that in several classes of problems an optimal design for estimating all the parameters is supported only on certain points of symmetry. This is applied to the problem (introduced by Scheffé (1958)) of multilinear regression on the simplex. In Section 3 we consider optimality when nuisance parameters are present. A new sufficient condition for optimality is given. A corrected version is given of the condition which Karlin and Studden (1966a) state as equivalent to optimality, and we prove the natural invariance theorem involving this condition. These results are applied to the problem of multilinear regression on the simplex when estimating only some of the parameters. Section 4 consists primarily of a number of bounds on the efficiency of designs; these are summarized at the beginning of that section.

1. Preliminaries.

Basic model. Let $f_1, \cdots, f_k$ be k continuous real-valued linearly independent functions, called the *regression functions*, on a compact space $\mathfrak{X}$. Let $\theta_1, \cdots, \theta_k$ be k unknown parameters. We shall write these as column vectors f and θ. For any x in $\mathfrak{X}$ we may observe a random variable $Y(x)$ with mean $\theta' f(x) = \Sigma\theta_i f_i(x)$. Throughout this paper primes will denote transposes. The variance σ^2 of $Y(x)$ may be known or unknown but it is fixed independent of x. The observed $Y(x_i)$ and $Y(x_j)$ are uncorrelated. We assign various values to x and make N observations in all. A *design* ξ is a probability measure on $\mathfrak{X}$. If ξ assigns to points probabilities which are all multiples of $1/N$, $\xi(x)$ will be the proportion of observations of Y at x.

Optimality criteria. A number of criteria for optimality of a design have been suggested. See Kiefer (1960, page 383 ff.) for a discussion of some of these. We consider only two.

The *information matrix* of a design ξ, denoted $M(\xi)$, has components defined by

$$m_{ij}(\xi) = \int f_i(x) f_j(x)\, d\xi(x).$$

We will sometimes write $M(\xi) = \int f(x) f'(x)\, d\xi(x)$. If $M(\xi)$ is nonsingular, so

Received 3 June 1968; revised 25 February 1969.

[1] The research was supported in part by the Office of Naval Research under contract number Nonr-401(50). This paper is a revised version of the author's PhD thesis submitted to Cornell University.

that all the components of θ are estimable under ξ, then $\sigma^2 N^{-1} M^{-1}(\xi)$ is the covariance matrix of the best linear estimator of θ. A design ξ^* is called *D-optimal* if it minimizes $\det M^{-1}(\xi)$.

Define the function

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

If $M(\xi)$ is nonsingular, the variance of the best linear estimator of $EY(x) = \theta' f(x)$ is $\sigma^2 N^{-1} d(x, \xi)$. A design ξ^* is called *G-optimal* if it minimizes $\max_{x \varepsilon \mathfrak{X}} d(x, \xi)$. We will simply write $\max_x d(x, \xi)$.

Kiefer and Wolfowitz (1960) proved that D-optimality and G-optimality are equivalent, and ξ is optimal if and only if $\max_x d(x, \xi) = k$.

Since $\int d(x, \xi)\, d\xi = k$, it is therefore immediate that an optimal ξ can have points of support only where $d(x, \xi) = k$.

The model discussed so far can be generalized as follows (See Kiefer (1961)). We write

$$\theta = \begin{pmatrix} \theta^{(1)} \\ \theta^{(2)} \end{pmatrix} \qquad f(x) = \begin{pmatrix} f^{(1)}(x) \\ f^{(2)}(x) \end{pmatrix}$$

where $\theta^{(1)}$ and $f^{(1)}$ are s-vectors and $\theta^{(2)}$ and $f^{(2)}$ are r-vectors, with $r = k - s$. As before $EY(x) = \theta' f(x)$, but this time we are interested in estimating only $\theta^{(1)}$ rather than all of θ. If $M(\xi)$ is nonsingular, so that all of θ is estimable under ξ, the inverse below can be written directly. If $\theta^{(2)}$ is not estimable, i.e., M_3 is singular, we understand that M^{-1} is a pseudo-inverse. (This is discussed in Chernoff (1953) and in Section 3 of this paper.) We write

$$M(\xi) = \begin{Vmatrix} M_1(\xi) & M_2(\xi) \\ M_2'(\xi) & M_3(\xi) \end{Vmatrix} \qquad M^{-1}(\xi) = \begin{Vmatrix} M^{(1)}(\xi) & M^{(2)}(\xi) \\ M^{(2)'}(\xi) & M^{(3)}(\xi) \end{Vmatrix}$$

where $M_1(\xi)$ and $M^{(1)}(\xi)$ are $s \times s$ matrices. The covariance matrix of the best linear estimator of $\theta^{(1)}$ is then $\sigma^2 N^{-1} M^{(1)}(\xi)$. A design ξ^* is called *D-optimal* for estimating s out of k parameters if it minimizes $\det M^{(1)}(\xi)$. We will sometimes write $M^*(\xi) = [M^{(1)}(\xi)]^{-1} = M_1(\xi) - M_2(\xi) M_3^{-1}(\xi) M_2'(\xi)$.

Although there is no natural optimality criterion analogous to G-optimality, if $M(\xi)$ is nonsingular we still define

$$\begin{aligned} d_{s|k}(x, \xi) &= f'(x) M^{-1}(\xi) f(x) - f^{(2)'}(x) M_3^{-1}(\xi) f^{(2)}(x) \\ &= (f^{(1)}(x) - M_2 M_3^{-1} f^{(2)}(x))' M^{(1)} (f^{(1)}(x) - M_2 M_3^{-1} f^{(2)}(x)) \end{aligned}$$

where in the second expression we have suppressed the ξ for greater legibility. Kiefer (1961) proved that when $M(\xi)$ is nonsingular, ξ is D-optimal for estimating $\theta^{(1)}$ if and only if $\max_x d_{s|k}(x, \xi) = s$.

If $M(\xi)$ is singular there is no simple known theorem analogous to this result. This is discussed in connection with Theorems 3.1 and 3.2 of this paper.

Invariance. We shall refer repeatedly to the result summarized here. Suppose G is a compact group of transformations on $\mathfrak{X}$. Let $\bar{G}$ be a group of linear transformations on the space of θ of the form

$$A = \begin{Vmatrix} B & 0 \\ 0 & C \end{Vmatrix}$$

where B is an $s \times s$ matrix of determinant ± 1. (When estimating all of θ, i.e., $s = k$, B is the entire $k \times k$ matrix A.) Assume there is a homomorphism from G to $\bar{G}$. (Note then that $A_{g_1 g_2} = A_{g_1} A_{g_2}$.) And suppose $\theta' f(x) = (A_g \theta)' f(gx)$.

A design ξ is called *G-invariant* if $\xi(B) = \xi(gB)$ for all g in G and Borel sets B. Then under the above conditions, there is a G-invariant optimal design. (See Kiefer (1959, page 296), (1960, page 387), and (1961, page 302).)

Number of points needed for optimality. When estimating θ, k-dimensional, a direct argument shows that there is an optimal design supported on at most $k(k+1)/2$ points. (See Kiefer (1960, page 389), or Farrell, Kiefer, and Walbran (1965, page 114 ff.).) An example is given in Section 4 for which all $k(k+1)/2$ points are necessary.

The lower bound on the number of points needed for optimality is clearly k, when estimating θ. We remark that if a design supported on k points is optimal, the design must be uniform on those points.

When estimating $\theta^{(1)}$ it can be shown that there is an optimal design supported on at most $s(s+1)/2 + rs$ points. (See Chernoff (1953, page 590 ff.), Stone (1959, page 68), or Kiefer (1961, page 303).) An example is given in Section 4 for which this bound is attained.

The minimum number of points needed to estimate $\theta^{(1)}$ (not necessarily optimally) may be any number from s to k. It is not hard to show that if $\theta^{(1)}$ is estimable using an s-point design then $f^{(2)}$ is zero on the support of such a design, and the best such design is uniform on s points. For let ξ_s be any s-point design which estimates $\theta^{(1)}$. There is a nonsingular matrix

$$L = \begin{Vmatrix} I & L_3 \\ 0 & L_2 \end{Vmatrix}$$

with

$$LM(\xi_s)L' = \begin{Vmatrix} M^*(\xi_s) & 0 & 0 \\ 0 & N_3 & 0 \\ 0 & 0 & 0 \end{Vmatrix}$$

where N_3 is a square nonsingular matrix of size $\geqq 0$. But rank $LM(\xi_s)\,L' =$ rank $M(\xi_s) \leqq s$. Since ξ_s estimates $\theta^{(1)}$, rank $M^*(\xi_s) = s$ and therefore N_3 must have size 0. So $f^{(2)}$ is 0 on the support of ξ_s and $M^*(\xi) = M_1(\xi)$ for any ξ supported on these points. The best such ξ is therefore uniform.

2. Optimal designs for estimating θ. When the k regression functions are the monomials of degree $\leqq k-1$ on an interval the optimal design is supported on k points, as described by Guest (1958) and Hoel (1958). If more than one variable is present, natural generalizations of the interval might be the cube, the ball and the simplex. Of these, the simplex seems to give the most analogous and the most simple results. Farrell, Kiefer and Walbran (1965) compare the spaces in some detail. If the regression functions are all the monomials of degree $\leqq n$ the problem is, in general, quite difficult. However if the regression functions belong to a more restricted class we can sometimes obtain extensive results.

We begin with three theorems in which symmetry of $\mathfrak{X}$ and restrictions on the regression functions enable us to make assertions about the support of the optimal designs. For the difficulty in proving analogues for estimation of $\theta^{(1)}$, see the lemma preceding Theorem 3.5.

Scheffé (1958, page 352) introduced the *special n-tic* polynomials, defined as the multilinear polynomials of total degree $\leqq n$. Under special n-tic regression we may take the regression functions $f_i(x)$ to be of the form $x_{i_1} \cdots x_{i_p}$, with $i_1 < \cdots < i_p$ and $p \leqq n$. If there are q independent variables then $p \geqq 0$; of course $n \leqq q$, and the number of regression functions is $k = \sum_{p=0}^{n} \binom{q}{p}$. If there are $q + 1$ variables constrained by $\sum x_i = \alpha$ then $p \geqq 1$, $n \leqq q + 1$, and the number of functions $k = \sum_{p=1}^{n} \binom{q+1}{p}$.

THEOREM 2.1. *Suppose the regression is special n-tic on a set $\mathfrak{X}$ in Euclidean q-space R^q. Then an optimal design ξ can have no points of support in the interior of any line segment in $\mathfrak{X}$ on which all the variables but one are constant.*

PROOF. Suppose ξ is optimal. Then $M^{-1}(\xi)$ exists and is positive definite and $d(x, \xi) = f'(x)M^{-1}(\xi)f(x) \geqq \lambda |f(x)|^2$, where $\lambda > 0$ is the smallest eigenvalue of $M^{-1}(\xi)$ and $|f(x)|$ is the Euclidean norm of $f(x)$. As $|x| \to \infty$ so does $|f(x)|$ and therefore so does $d(x, \xi)$. Since $f(x)$ is multilinear $d(x, \xi)$ is at most quadratic in each variable. Now hold all the variables but one constant. The restriction of $d(x, \xi)$ to this line is a non-negative unbounded quadratic function. Therefore it is strictly convex and cannot have a maximum in the interior of any segment of the line.

For $f(x)$ as above we list two corollaries.

Special cases of $\mathfrak{X}$.

(1) $\mathfrak{X}$ is an arbitrary set in R^q.

Any optimal design is supported on the boundary of $\mathfrak{X}$.

(2) $\mathfrak{X}$ is a set whose convex hull is the q-cube.

Any optimal design is supported on the vertices of the cube. If $n = q$ the unique optimal design is uniform on the vertices (because the number of regression functions equals the number of points of support.) If $n < q$, the design uniform on the vertices is the unique optimal design which is invariant under the group of symmetries of the cube.

THEOREM 2.2. *Suppose the regression is special n-tic on a set $\mathfrak{X}$ in R^{q+1} satisfying the linear constraint $\sum x_i = \alpha$ and symmetric under the interchange of some two coordinates. Let I be any line segment in $\mathfrak{X}$ which is invariant under the interchange of those same two coordinates, and on which all coordinates but those two are constant. Then an optimal design ξ can have points of support on I only at the midpoint and at the end points.*

PROOF. Let ξ be optimal. As before $d(x, \xi) \to \infty$ as $|x| \to \infty$ and $d(x, \xi)$ is at most quadratic in each variable. Without loss of generality $\mathfrak{X}$ and I are symmetric under interchange of x_1 and x_2. The restriction of $d(x, \xi)$ to the line containing I, when expressed in terms of one variable, say x_1, is at most quartic. Since $d(x, \xi)$ is non-negative and not constant on the line, it can have at most one interior maximum in I.

Now if π is the permutation which just interchanges x_1 and x_2 there is a linear mapping L so that $Lf(x) = f(\pi x)$. By the standard invariance results (see Section 1) there is an optimal design ξ' which is invariant under π. Since ξ' is optimal, $M(\xi) = M(\xi')$. (See Kiefer and Wolfowitz (1960).) It follows that $d(x, \xi) = d(x, \xi') = d(\pi x, \xi') = d(\pi x, \xi)$. That is, $d(x, \xi)$ is symmetric, and thus can have maxima in I only at the midpoint and end points.

Special cases of $\mathfrak{X}$.

(1) $\mathfrak{X}$ is any convex body symmetric under all permutations of the coordinates (e.g., q-ball), constrained by $\sum x_i = \alpha$.

Any optimal design is supported on the boundary of $\mathfrak{X}$ and the center point of $\mathfrak{X}$.

(2) $\mathfrak{X}$ is the q-simplex.

Any optimal design is supported on the barycenters. If $n = q + 1$ the unique optimal design is uniform on the barycenters (because the number of functions equals the number of points of support).

(3) $\mathfrak{X}$ is determined by $0 \leqq x_i \leqq 1$, $\sum x_i = \alpha$, $1 < \alpha < q + 1$.

Any optimal design is supported on points of the form $(0, \cdots, 0, 1, \cdots, 1, \beta, \cdots, \beta)$ with $\beta = (\alpha - N_1)/(q + 1 - N_0 - N_1)$, where N_0 is the number of 0's and N_1 the number of 1's. This space is considered in a different problem by Keilson (1966).

We remark that 2 and 3 are special cases of 1.

For the last theorem of this type we assume that the regression functions are the monomials in q independent variables which involve at most n variables, and which in each variable are at most quadratic (e.g., $1, x, x^2, xy, x^2y, x^2y^2z$, etc.). Of course $n \leqq q$. The number of such functions is $\sum_{i=0}^{n} \binom{q}{i} 2^i$.

THEOREM 2.3. *Let the regression functions be as described above, and let* $\mathfrak{X}$ *be any space in* R^q *which for some coordinate* x_i *is symmetric under interchange of* x_i *and* $-x_i$. *Let* I *be a line segment in* $\mathfrak{X}$ *on which all coordinates but* x_i *are constant and which is symmetric under interchange of* x_i *and* $-x_i$. *Then an optimal design* ξ *can have points of support on* I *only at the midpoint and at the end points.*

This proof is similar to the previous two.

Special cases of $\mathfrak{X}$.

(1) $\mathfrak{X}$ is any convex body which for each x_i is symmetric under interchange of x_i and $-x_i$ (e.g. q-ball).

Any optimal design is supported on the boundary and the origin. This result does not overlap the first special case of Theorem 2.2 because the functions are different in the two cases.

(2) $\mathfrak{X}$ is the q-cube.

Any optimal design is supported on the lattice of points with coordinates only 0 or ± 1. (There are 3^q such points.) If $n = q$ the unique optimal design is uniform on this lattice.

Again we note that 2 is a special case of 1.

We conclude this section by considering the example of special n-tic regression on the q-simplex.

The design ξ^0, which was introduced by Scheffé (1958), assigns equal measure to the points on the simplex $(1, 0, \cdots, 0)$, $(\frac{1}{2}, \frac{1}{2}, 0, \cdots, 0)$, $\cdots$, $(1/n, \cdots, 1/n, 0, \cdots, 0)$ and their images under permutations of the coordinates. These will sometimes be referred to as the barycenters of depth $\leqq n$. The number of such points is $\sum_{p=1}^{n} \binom{q+1}{p} = k$, the number of regression functions.

Trivially ξ^0 is optimal for any q when $n = 1$. It was shown by Kiefer (1961, p. 320) that ξ^0 is optimal for all q when $n = 2$, and by Uranisi (1964) that ξ^0 is optimal for all q when $n = 3$. The natural conjecture for greater n is shown here to be only partly true.

THEOREM 2.4. *For special n-tic regression on the q-simplex,* $4 \leqq n \leqq q + 1$, *the design* ξ^0 *is optimal when* $n = q + 1$ *and is not optimal when* $n < q + 1$.

PROOF. Optimality when $n = q + 1$ is just the second special case of Theorem 2.2. We now prove that ξ^0 is not optimal when $4 \leqq n < q + 1$.

First observe that if L is a nonsingular matrix, $g(x) = Lf(x)$, and $N(\xi) = \int g(x)g'(x)\,d\xi(x)$, then ξ^* maximizes $\det N(\xi)$ if and only if it maximizes $\det M(\xi)$ and $g'(x)N^{-1}(\xi)g(x) = f'(x)M^{-1}(\xi)f(x)$ identically in x and ξ. The following k functions $g_i(x)$ are so obtained.

$$(2.1)\quad p^p x_{i_1} \cdots x_{i_p}[a_{0,p} + a_{1,p} \sum x_{j_1} + a_{2,p} \sum x_{j_1}x_{j_2} + \cdots + a_{n-p,p} \sum x_{j_1} \cdots x_{j_{n-p}}].$$

Here $1 \leqq p \leqq n$, $a_{s,p} = (-1)^s p^{s-1}(p + s)$, $1 \leqq i_1 < \cdots < i_p \leqq q + 1$, and in each summation $1 \leqq j_1 < j_2 < j_3 < \cdots \leqq q + 1$ and none of the j's are in $\{i_1, \cdots, i_p\}$.

It is immediate that such a polynomial equals 1 at the barycenter which has all coordinates zero except for $i_1, \cdots, i_p$, and that the polynomial equals 0 at all other barycenters of depth $\leqq p$. At a barycenter of depth $p + m \leqq n$, $m > 0$, either $x_{i_1} \cdots x_{i_p} = 0$ or the expression in brackets is equal to

$$(2.2)\quad \sum_{s=0}^{m} (-1)^s p^{s-1}(p + s) \binom{m}{s} (p + m)^{-s} = (1 - p/(p + m))^m - \frac{d}{dx} \sum_{s=0}^{m} \binom{m}{s} x^s (p + m)^{-s} \Big|_{x=-p}$$

which equals 0. Therefore the polynomials are orthogonal with respect to ξ^0 and the sum of their squares equals $k^{-1} d(x, \xi^0)$.

Because $n < q + 1$ there is a barycenter of depth $p + m = n + 1$. Let z be such a barycenter, and consider $i_1, \cdots, i_p$ such that $x_{i_1} \cdots x_{i_p} \neq 0$ at z. Evaluated there the expression in brackets in (2.1) equals

$$\sum_{s=0}^{m-1} (-1)^s p^{s-1} (p + s) \binom{m}{s} (p + m)^{-s}$$
$$= 0 - (-1)^m p^{m-1} (p + m)^{-(m-1)} = (-p)^{(n-p)} (n + 1)^{-(n-p)} \quad \text{using (2.2).}$$

Therefore at z

$$k^{-1} d(z, \xi^0) = \sum_{i=1}^{k} g_i^2(z) = \sum_{p=1}^{n} \binom{n+1}{p} (p/(n + 1))^{2n}.$$

As a lower bound on this we consider the upper two terms,

$$(n + 1)(n/(n + 1))^{2n} + \tfrac{1}{2}n(n + 1)[(n - 1)/(n + 1)]^{2n}$$

which we write as $A + B$. When $n = 4$ a direct computation yields $A + B = 78658/78125 > 1$. For $n \geqq 4$ we show as follows that $A + B$ is monotone increasing (in fact it increases without limit).

Treating n as a continuous parameter,

$$\frac{d}{dn} \log A = 3/(n + 1) - 2 \log (1 + 1/n) > 3/(n + 1) - 2/n.$$

Therefore A is monotone increasing for $n \geqq 2$. In fact

$$A = (n + 1)(1 + 1/n)^{-2n} \sim (n + 1)e^{-2}$$

so $A \rightarrow \infty$.

Likewise,

$$\frac{d}{dn} \log B = [6n^2 - n - 1]/[n(n - 1)(n + 1)] - 2 \log (1 + 2/(n - 1))$$
$$> [6n^2 - n - 1]/[n(n - 1)(n + 1)] - 4/(n - 1)$$
$$= [2n^2 - 5n - 1]/[n(n - 1)(n + 1)].$$

The numerator of the last expression has zeros $(5 \pm 33^{\frac{1}{2}})/4$, so B is monotone increasing for $n \geqq 3$. Therefore $d(z, \xi^0) > k$ for $n \geqq 4$, completing the proof.

An optimal design when $n < q + 1$ is not known. As has been shown in the second special case of Theorem 2.2 any optimal design must be supported on the barycenters. The fact that $d(z, \xi^0) > k$ when z is a barycenter of depth $n + 1$ suggests that an optimal design would assign positive measure to these points, but it is not known whether this is correct, or what weights should be used.

3. Optimal designs for estimating $\theta^{(1)}$. In this section we first prove some general results for estimating s out of k parameters and then apply them in the case of special n-tic regression on the q-simplex.

As always we assume below that the regression functions f_i are continuous and that $\mathfrak{X}$ is compact. The first result is a sufficient condition for optimality.

We first consider more precisely the case when $M(\xi)$ is singular. Chernoff (1953) gives the following definition of a pseudo-inverse. For X a symmetric nonnegative definite matrix, let Y be any symmetric matrix such that $X + \lambda Y$ is postive definite for sufficiently small $\lambda > 0$. Define the pseudo-inverse $X^{-1} = \lim_{\lambda \to 0^+} (X + \lambda Y)^{-1}$. Write the ijth element of X^{-1} as x^{ij}. Chernoff proves that the diagonal elements of X^{-1} are independent of the particular Y used, and if x^{ii} and x^{jj} are finite then x^{ij} is finite and independent of Y. X^{-1} may have infinite entries, but we will only be interested in $M^{(1)}(\xi)$, which will always be finite and well-defined if $\theta^{(1)}$ is estimable under ξ.

If for designs ξ and ξ' we define $\xi_\epsilon = (1 - \epsilon)\xi + \epsilon\xi'$, and if $M(\xi_\epsilon)$ is nonsingular for $0 < \epsilon < 1$, then this pseudo-inverse $M^{-1}(\xi) = \lim_{\epsilon \to 0^+} M^{-1}(\xi_\epsilon)$, $M^{(1)}(\xi) = \lim_\epsilon M^{(1)}(\xi_\epsilon)$, and $M^*(\xi) = \lim_\epsilon M^*(\xi_\epsilon)$. (As so defined $\sigma^2 N^{-1} M^{(1)}(\xi)$ really is the covariance matrix of the best linear estimator of $\theta^{(1)}$ under ξ.)

THEOREM 3.1. *Let ξ^* be any design and ξ' any design such that $M(\xi_\epsilon)$ is nonsingular for $0 < \epsilon < 1$, where $\xi_\epsilon = (1 - \epsilon)\xi^* + \epsilon\xi'$. Then the statements* (3.1) *and* (3.2) *are equivalent, and either implies* (3.3).

(3.1) $\lim_{\epsilon \to 0^+} d_{s \mid k}(x, \xi_\epsilon)$ *exists, and* $\max_{x \varepsilon \mathfrak{X}} \lim_{\epsilon \to 0^+} d_{s \mid k}(x, \xi_\epsilon) = s$.

(3.2) $\lim_{\epsilon \to 0^+} \max_{x \varepsilon \mathfrak{X}} d_{s \mid k}(x, \xi_\epsilon)$ *exists and* $= s$.

(3.3) ξ^* *is D-optimal for estimating* $\theta^{(1)}$.

PROOF. We prove first that (3.2) implies (3.1).

If $\lim_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon) = s$ there is some $\epsilon_0 > 0$ with $\max_x d_{s \mid k}(x, \xi_\epsilon) \leqq s + 1$ for $\epsilon \leqq \epsilon_0$. Therefore $0 \leqq d_{s \mid k}(x, \xi_\epsilon) \leqq s + 1$ for $x \varepsilon \mathfrak{X}$, $0 < \epsilon \leqq \epsilon_0$. For each x, $d_{s \mid k}(x, \xi_\epsilon)$ is a rational function in ϵ. It is bounded as $\epsilon \to 0$, and therefore $\lim_\epsilon d_{s \mid k}(x, \xi_\epsilon)$ exists.

We now want to show that for $\delta > 0$ and $\epsilon < \epsilon_0$, if $|x - x'|$ is sufficiently small then

$$(3.4) \qquad |d_{s \mid k}(x, \xi_\epsilon) - d_{s \mid k}(x', \xi_\epsilon)| < \delta.$$

Write

$$d_{s \mid k}(x, \xi_\epsilon) = \textstyle\sum_{i,j=1}^k a_{ij}(\epsilon) f_i(x) f_j(x) = \sum_{i=1}^N b_i(\epsilon) F_i(x),$$

where the F_i's are linear combinations of the terms $f_j f_l$, chosen to be linearly independent. Then there is a set $\{x_1, \cdots, x_N\}$ so that F, the matrix with entries $F_i(x_j)$, is nonsingular. Then (letting $|\;|$ denote the Euclidean norm of a vector and the corresponding operator norm of a matrix)

$$|b(\epsilon)| = |b(\epsilon)' F F^{-1}| \leqq |b(\epsilon)' F| \, |F^{-1}|$$

which is bounded as $\epsilon \to 0$ because $b(\epsilon)' F(x) = d_{s \mid k}(x, \xi_\epsilon)$ is bounded. Therefore for $0 < \epsilon < \epsilon_0$ and $|x - x'|$ sufficiently small

$$|d_{s \mid k}(x, \xi_\epsilon) - d_{s \mid k}(x', \xi_\epsilon)| \leqq N \max_i \sup_{0 < \epsilon < \epsilon_0} |b_i(\epsilon)| \max_j |F_j(x) - F_j(x')| < \delta$$

proving (3.4). The last inequality uses the continuity of the F_i's and the compactness of $\mathfrak{X}$.

Let x_ϵ be a point at which $\max_x d_{s \mid k}(x, \xi_\epsilon)$ is attained. Since $\mathfrak{X}$ is compact there is a point x_0 and a sequence x_{ϵ_n} approaching x_0 as $\epsilon_n \to 0$. Pick $\epsilon_1 \leqq \epsilon_0$ so that if

$0 < \epsilon_n < \epsilon_1$ then

$$|s - d_{s \mid k}(x_{\epsilon_n}, \xi_{\epsilon_n})| < \delta$$

and $|x_{\epsilon_n} - x_0|$ is small enough so that

$$|d_{s \mid k}(x_{\epsilon_n}, \xi_{\epsilon_n}) - d_{s \mid k}(x_0, \xi_{\epsilon_n})| < \delta$$

and

$$|d_{s \mid k}(x_0, \xi_{\epsilon_n}) - \lim_{\epsilon_n} d_{s \mid k}(x_0, \xi_{\epsilon_n})| < \delta.$$

(Recall that we have shown that $\lim_\epsilon d_{s \mid k}(x, \xi_\epsilon)$ exists for each x.) Therefore

$$|s - \lim_\epsilon d_{s \mid k}(x_0, \xi_\epsilon)| < 3\delta, \quad \text{for all } \delta > 0.$$

So $\max_x \lim_\epsilon d_{s \mid k}(x, \xi_\epsilon) \geqq s$. But also

$$\max_x \lim_\epsilon d_{s \mid k}(x, \xi_\epsilon) \leqq \lim_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon) = s.$$

Therefore $\max_x \lim_\epsilon d_{s \mid k}(x, \xi_\epsilon) = s$.

We next prove that (3.1) implies (3.2).

We assume that $\lim_\epsilon d_{s \mid k}(x, \xi_\epsilon)$ exists for all x and is $\leqq s$. Just as before we obtain that for $\epsilon <$ some ϵ_0 and for $|x - x'|$ sufficiently small, (3.4) holds. Now choose a sequence ϵ_n so that

$$\max_x d_{s \mid k}(x, \xi_{\epsilon_n}) \to \limsup_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon).$$

The value $\max_x d_{s \mid k}(x, \xi_{\epsilon_n})$ is attained at x_{ϵ_n}. A subsequence (again written x_{ϵ_n}) approaches some x_0. So choose $\epsilon_1 \leqq \epsilon_0$ so that if $0 < \epsilon_n < \epsilon_1$ then

$$|\limsup_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon) - d_{s \mid k}(x_{\epsilon_n}, \xi_{\epsilon_n})| < \delta$$

and

$$|d_{s \mid k}(x_{\epsilon_n}, \xi_{\epsilon_n}) - d_{s \mid k}(x_0, \xi_{\epsilon_n})| < \delta$$

and

$$|d_{s \mid k}(x_0, \xi_{\epsilon_n}) - \lim_\epsilon d_{s \mid k}(x_0, \xi_\epsilon)| < \delta.$$

Then, since δ is arbitrary,

$$\limsup_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon) = \lim_\epsilon d_{s \mid k}(x_0, \xi_\epsilon) \leqq s.$$

But for any ξ with $M(\xi)$ nonsingular, $\max_x d_{s \mid k}(x, \xi) \geqq s$, so

$$\liminf_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon) \geqq s,$$

proving that $\lim_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon)$ exists and equals s.

Finally we must show that (3.2) implies (3.3). But this is immediate from the corollary to the first part of Theorem 4.3, which we will prove easily from first principles. This completes the proof of Theorem 3.1.

We will show in Example 3.1 that (3.3) does not imply (3.2). However since that example illustrates several other things as well, we postpone it until we can give a full discussion.

Let us compare Theorem 3.1 with previously known results. Kiefer (1961) defines functions $\bar{D}(x, \xi^*)$, $D(\xi, \xi^*)$ and $D(x, \xi^*)$, and shows that for estimating $\theta^{(1)}$

$$\max_x \bar{D}(x, \xi^*) = s \Rightarrow \max_\xi D(\xi, \xi^*) = s \Leftrightarrow \xi^* \text{ optimal} \Rightarrow \max_x D(x, \xi^*) = s.$$

Later in this section we will consider the example of special $(q + 1)$-tic regression on the q-simplex, where $f^{(1)}$ consists of the multilinear monomials of degree $\leqq m$. In that example Kiefer's results are inconclusive, because for the design ξ^0 under consideration, $\max_x \bar{D}(x, \xi^0) > s$, $\max_x D(x, \xi^0) = s$, and $D(\xi, \xi^0)$ is extremely hard to compute. However Theorem 3.1 can be applied to show that ξ^0 is optimal.

Karlin and Studden (1966a, Section 6) and (1966b, Chapter 10), take the following approach. For any $r \times s$ matrix X define

$$d_s(x, \xi, X) = (f^{(1)}(x) - X'f^{(2)}(x))'[M^*(\xi)]^{-1}(f^{(1)}(x) - X'f^{(2)}(x)).$$

For any ξ let $X(\xi)$ be an $r \times s$ matrix satisfying

$$M_3(\xi)X(\xi) = M_2'(\xi). \tag{3.5}$$

Such an $X(\xi)$ is shown always to exist. If $M(\xi)$ is nonsingular, $X(\xi) = M_3^{-1}(\xi)M_2'(\xi)$, and $d_s(x, \xi, X(\xi))$ is equal to $d_{s\,|\,k}(x, \xi)$. If $M_3(\xi)$ is singular, $X(\xi)$ is not unique, but in any case the authors write $d_s(x, \xi, X(\xi)) = d_s(x, \xi)$, which is supposed to be well defined. Example 3.1 will show that neither $d_s(x, \xi)$ nor $\max_x d_s(x, \xi)$ is independent of the particular $X(\xi)$ chosen. It is then asserted that if $X(\xi)$ satisfies (3.5) for all ξ, then (i), (ii) and (iii) are equivalent.

(i) ξ^* maximizes $M^*(\xi)$. (D-optimality)

(ii) ξ^* minimizes $\max_x d_s(x, \xi, X(\xi))$.

(iii) $\max_x d_s(x, \xi^*, X(\xi^*)) = s$.

The proofs that (iii) implies (i) and (iii) implies (ii) are correct no matter what $X(\xi)$ is chosen. It is not clear that (ii) implies (iii) unless $X(\xi)$ is chosen in a certain way. Finally (i) implies (iii) only if $X(\xi)$ is chosen in a special way. Example 3.1 will show that (i) may hold while (iii) does not.

Before discussing the proof we point out that the theorem as stated does give a correct sufficient condition for optimality, which we state as a corollary.

Corollary. (Karlin and Studden) *If for some $r \times s$ matrix X $\max_x d_s(x, \xi, X) = s$ then ξ is D-optimal.*

Proof. It is not hard to show (see equation (6.9) of Karlin and Studden (1966a)) that $d_s(x, \xi, X)$ satisfies

$$\int d_s(x, \xi, X)\, d\xi(x) \geqq s, \tag{3.6}$$

with equality holding if and only if $M_3(\xi)X = M_2'(\xi)$. Thus if

$$\max_x d_s(x, \xi, X) = s,$$

equality holds in (3.6) and $M_3(\xi)X = M_2'(\xi)$. The correct portion of the theorem then gives the result.

We now consider the proof of the theorem.

The authors define a sequence of games. By taking the limit of the corresponding minimax strategies they obtain a certain matrix D_0.

In proving that (i) implies (iii) it is shown that if ξ^* satisfies (i) then

$$M_3(\xi^*)D_0 = M_2'(\xi^*). \tag{3.7}$$

Hence it is claimed that

$$\int d_s(x, \xi^*, D_0)\, d\xi(x) = \int d_s(x, \xi^*, X(\xi^*))\, d\xi(x).$$

This is true if we choose $X(\xi^*) = D_0$, but it is not true in general, as Example 3.1 will show.

There is certainly some ξ^* satisfying (i). For this ξ^*, (3.7) holds, so D_0 is an allowed choice for $X(\xi^*)$. In proving that (ii) implies (iii) it is claimed that

$$\min_\xi \max_\eta \int d_s(x, \xi, X(\xi))\, d\eta(x) \leqq \max_\eta \int d_s(x, \xi^*, D_0)\, d\eta(x).$$

It is not clear without additional argument that the inequality is true unless we choose $X(\xi^*) = D_0$.

Thus it is apparent how to define $X(\xi)$ to make the theorem true. Find D_0 and choose $X(\xi) = D_0$ whenever $M_3(\xi)D_0 = M_2'(\xi)$. Otherwise choose any $X(\xi)$ satisfying (3.5). Then the proof holds as it is, and the modified theorem is correct.

Alternatively we could replace $X(\xi)$ by D_0 for all ξ. We obtain this modification of the theorem if we show that $\max_x d_s(x, \xi, D_0) \geqq s$ for all ξ, with equality if and only if ξ is optimal. But this is quick to show. By (3.6) we have $\max_x d_s(x, \xi, D_0) \geqq s$, for all ξ. If equality holds then by the corollary given above, ξ is optimal. On the other hand if ξ^* is optimal then (3.7) holds, so choosing $X(\xi)$ as in the last paragraph we have $X(\xi^*) = D_0$ and

$$\max_x d_s(x, \xi^*, D_0) = \max_x d_s(x, \xi^*, X(\xi^*)) = s.$$

We write out the theorem just obtained.

THEOREM 3.2. (Karlin and Studden) *There is an* $r \times s$ *matrix* D_0 *so that the sets of designs* ξ^* *satisfying* (i), (ii), *or* (iii) *coincide.*

(i) ξ^* *maximizes* $\det M^*(\xi)$. (D-optimality)

(ii) ξ^* *minimizes* $\max_x d_s(x, \xi, D_0)$.

(iii) $\max_x d_s(x, \xi^*, D_0) = s$.

We will use this theorem later in this section. However since D_0 is in general so hard to determine, it seems that the corollary given earlier will usually be more applicable in examples than Theorem 3.2.

Both the Karlin-Studden Corollary and Theorem 3.1 give sufficient conditions for optimality. It seems as easy in principle to guess an X which works in the corollary as to guess a ξ' (hence ξ_ϵ) which works in Theorem 3.1. However after proving Theorem 3.7 by using a ξ_ϵ one can note that it appears hard to find an X with $\max_x d_s(x, \xi^0, X) = s$. In particular the matrix $X(\xi_\epsilon)$, corresponding to the ξ_ϵ used in proving Theorem 3.7, is uniquely determined and equal to $M_3^{-1}(\xi_\epsilon)M_2'(\xi_\epsilon) = 0$, but we cannot take $X = \lim_\epsilon X(\xi_\epsilon)$, since for that X, $\max_x d_s(x, \xi^0, X) > s$. On the other hand there are problems in which the X

method is very quick. Also we know that a suitable X always exists, D_0. We have not proved that a ξ_ϵ satisfying (3.1) and (3.2) always exists.

Let us conclude this discussion with the promised example.

EXAMPLE 3.1. Let $\mathfrak{X}$ be the unit interval, $0 \leqq x \leqq 1$, let $s = 1$ and $k = 2$, let $f_1(x) = x$, $f_2(x) = 1 - x$, and let ξ^* be concentrated at 1.

If ξ' is concentrated at $c/(c+1)$, for $c \geqq 0$, we have

$$M(\xi_\epsilon) = \epsilon(c+1)^{-2} \begin{Vmatrix} (c+1)^2(1-\epsilon)\epsilon^{-1} + c^2 & c \\ c & 1 \end{Vmatrix}.$$

Then $d_{s \mid k}(x, \xi_\epsilon) = (1-\epsilon)^{-1}(x - c(1-x))^2$. This is maximized at an end point of the interval, and $\max_x d_{s \mid k}(x, \xi_\epsilon) = (1-\epsilon)^{-1} \max(1, c^2)$. If $c \leqq 1$,

$$\lim_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon) = 1 = s,$$

proving that ξ^* is optimal, by Theorem 3.1. If $c > 1$,

$$\lim_\epsilon \max_x d_{s \mid k}(x, \xi_\epsilon) = c^2 > s,$$

demonstrating that (3.3) does not imply (3.2).

Let us now use the Karlin-Studden method. We have

$$M(\xi^*) = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix},$$

so any 1×1 matrix X satisfies $M_3(\xi^*)X = M_2'(\xi^*)$. If we let $X = c$, $-\infty < c < \infty$, then $d_s(x, \xi^*, X) = (x - c(1-x))^2$. This has maximum value $\max(1, c^2)$, which equals $1 = s$ if $|c| \leqq 1$, and equals $c^2 > s$ if $|c| > 1$. Thus in the theorem as originally stated (i) may hold while (iii) does not. In this example D_0 can be computed and found to equal 0. Thus for any design ξ,

$$\max_x d_s(x, \xi, D_0) = \max_x x[M^*(\xi)]^{-1}x = [M^*(\xi)]^{-1},$$

and a design minimizes $\max_x d_s(x, \xi, D_0)$ if and only if it maximizes $\det M^*(\xi)$.

Finally we remark that if $X = -1$, then $d_s(x, \xi^*, X) = 1$, identically in x. The fact that $d_s(x, \xi^*, X)$ can conceivably be constant on certain intervals will cause considerable difficulty in a later example. (See the lemma preceding Theorem 3.5.)

Before going further we point out a fact which will be used from time to time. If L is a nonsingular $k \times k$ matrix of the form

$$L = \begin{Vmatrix} L_1 & L_2 \\ 0 & L_3 \end{Vmatrix} \tag{3.8}$$

with L_1 $s \times s$, and if $LML' = N$, then $N^* = L_1M^*L_1'$. If $Lf = g$ and $M(\xi)$ is nonsingular, then $d_{s \mid k}(x, \xi)$ is the same whether written in terms of f and M or g and N.

Let us now consider the question of invariance in connection with Theorem 3.2. When estimating all k parameters, under certain assumptions we had the case that for ξ^* optimal, $d(x, \xi^*)$ was invariant under some group of transformations. The next theorem gives an analogous result.

THEOREM 3.3. *Let* G, $\bar{G}$ *and the matrices*

$$A_g = \begin{Vmatrix} B_g & 0 \\ 0 & C_g \end{Vmatrix}$$

be as described in the section on invariance in Section 1. *There exists an* $r \times s$ *matrix* D *such that for any* A_g *in* $\bar{G}$ *we have* $C_g^{-1}DB_g = D$, *and if* ξ^* *is optimal then* $d_s(x, \xi^*, D) = d_s(gx, \xi^*, D)$ *for any* g *in* G *and* x *in* $\mathfrak{X}$. *Moreover the sets of designs* ξ^* *satisfying* (i), (ii) *or* (iii) *coincide.*

(i) ξ^* *maximizes* $\det M^*(\xi)$. (D-optimality)

(ii) ξ^* *minimizes* $\max_x d_s(x, \xi, D)$.

(iii) $\max_x d_s(x, \xi^*, D) = s$.

PROOF. First we must define D. Let D_0 be as in Theorem 3.2. For g in G and corresponding matrix A_g, write $C_g^{-1}D_0B_g = D_g$. Let μ denote Haar measure on G with $\mu(G) = 1$. We define $D = \int D_g \, d\mu(g)$.

For any g_1 in G we have

$$C_{g_1}^{-1}DB_{g_1} = \int C_{g_1}^{-1}C_g^{-1}D_0B_gB_{g_1} \, d\mu(g) = \int C_{gg_1}^{-1}D_0B_{gg_1} \, d\mu(g)$$

because $A_gA_{g_1} = A_{gg_1}$, as mentioned in Section 1.

But this last expression equals

$$\int C_{g'}^{-1}D_0B_{g'} \, d\mu(g'g_1^{-1}) = \int C_{g'}^{-1}D_0B_{g'} \, d\mu(g') = D$$

by the invariance of Haar measure. Thus D is invariant in the sense defined.

Now suppose ξ^* is optimal, i.e., (i) holds. We must show that $d_s(x, \xi^*, D)$ is invariant.

For each fixed g, note first that $\theta'f(x) = (A_g\theta)'f(gx) = \theta'A_g'f(gx)$ for all θ and x, and therefore $f(gx) = A_g'^{-1}f(x)$. Now for any $r \times s$ matrix X, we have

$$\begin{aligned} d_s(gx, \xi^*, X) &= f'(gx)(I, -X')'[M^*(\xi^*)]^{-1}(I, -X')f(gx) \\ &= f'(x)(I, -B_g'X'C_g'^{-1})'B_g^{-1}[M^*(\xi^*)]^{-1}B_g'^{-1}(I, -B_g'X'C_g'^{-1})f(x). \end{aligned}$$

Let ξ_0 be any invariant optimal design. Then for each g,

$$A_g'M(\xi_0)A_g = \int f(g^{-1}x)f'(g^{-1}x) \, d\xi_0(x) = M(\xi_0),$$

and hence $B_g'M^*(\xi_0)B_g = N^*(\xi_0)$, by the remarks in connection with (3.8). Since $M^*(\xi)$ is the same for all optimal ξ (Kiefer (1961)), we conclude that $B_g'M^*(\xi^*)B_g = M^*(\xi^*)$. Therefore

$$\begin{aligned} (3.9) \quad d_s(gx, \xi^*, X) &= f'(x)(I, -B_g'X'C_g'^{-1})'[M^*(\xi^*)]^{-1}(I, -B_g'X'C_g'^{-1})f(x) \\ &= d_s(x, \xi^*, C_g^{-1}XB_g). \end{aligned}$$

In particular, $d_s(gx, \xi^*, D) = d_s(x, \xi^*, D)$.

We now must show that (i), (ii), and (iii) are equivalent. First we prove that (i) implies (iii). Let ξ^* be optimal. Then by Theorem 3.2, we know that $d_s(x, \xi^*, D_0) \leq s$ for all x. Therefore, by (3.9),

$$d_s(x, \xi^*, D_g) = d_s(gx, \xi^*, C_{g^{-1}}^{-1}C_g^{-1}D_0B_gB_{g^{-1}}) = d_s(gx, \xi^*, D_0) \leq s$$

for all x.

Observe that if P is any symmetric positive definite $s \times s$ matrix and z any s-vector, then $z'Pz$ is convex in z. For fixed x let $z_g = (I, -D_g')f(x)$. Then by Jensen's inequality we have

$$\begin{aligned} d_s(x, \xi^*, D) &= f'(x)(I, -\textstyle\int D_g'\, d\mu(g))'[M^*(\xi^*)]^{-1}(I, -\textstyle\int D_g'\, d\mu(g))f(x) \\ &= \textstyle\int [f'(x)(I, -D_g')']\, d\mu(g)[M^*(\xi^*)]^{-1} \int [(I, -D_g')f(x)]\, d\mu(g) \\ &\leqq \textstyle\int [f'(x)(I, -D_g')'[M^*(\xi^*)]^{-1}(I, -D_g')f(x)]\, d\mu(g) \\ &= \textstyle\int d_s(x, \xi^*, D_g)\, d\mu(g) \\ &\leqq s. \end{aligned}$$

To see that equality holds for some x, recall that (3.6) says that

$$\int d_s(x, \xi, X)\, d\xi(x) \geqq s,$$

for any $r \times s$ matrix X. Thus (i) implies (iii).

Moreover from (3.6) we see that $\max_x d_s(x, \xi, D) \geqq s$ for any ξ, so (iii) implies (ii). But there is a ξ, namely an optimal ξ, for which $\max_x d_x(x, \xi, D) = s$, so (ii) implies (iii). Finally, (iii) implies (i) by the corollary to the theorem of Karlin and Studden given earlier. This completes the proof of the theorem.

The remaining results in this section concern special situations.

THEOREM 3.4. *If ξ_s is optimal for estimating the s out of s coefficients of $Af^{(1)} + Bf^{(2)}$, where A is $s \times s$ nonsingular and B is $s \times r$, and if $f^{(2)} = 0$ on the support of ξ_s, then ξ_s is optimal for estimating $\theta^{(1)}$* out of θ.

REMARK. As a special case we may have $A = I$ and $B = 0$.

PROOF. Write $g^{(1)} = Af^{(1)} + Bf^{(2)}$. There is a nonsingular $s \times s$ matrix L_1 with $L_1[\int g^{(1)}g^{(1)'}\, d\xi_s]L_1' = I$. Write $h^{(1)} = L_1 g^{(1)} = L_1Af^{(1)} + L_1Bf^{(2)}$. Since ξ_s is optimal for the s out of s coefficients of $g^{(1)}$,

$$s \geqq g^{(1)'}(x)[\textstyle\int g^{(1)}g^{(1)'}\, d\xi_s]^{-1}g^{(1)}(x) = h^{(1)'}(x)h^{(1)}(x). \tag{3.10}$$

Now let

$$L = \begin{Vmatrix} L_1A & L_1B \\ 0 & I \end{Vmatrix} \quad \text{and} \quad Lf = \begin{pmatrix} h^{(1)} \\ f^{(2)} \end{pmatrix} = h.$$

Let ξ' be a design so that $M(\xi_\epsilon)$ is nonsingular for $0 < \epsilon < 1$, where $\xi_\epsilon = (1 - \epsilon)\xi_s + \epsilon\xi'$. Then

$$\begin{aligned} d_{s\mid k}(x, \xi_\epsilon) &= f'(x)M^{-1}(\xi_\epsilon)f(x) - f^{(2)'}(x)M_3^{-1}(\xi_\epsilon)f^{(2)}(x) \\ &= h'(x)[(1-\epsilon)LM(\xi_s)L' + \epsilon LM(\xi')L']^{-1}h(x) \\ &\quad - f^{(2)}(x)M_3^{-1}(\xi_\epsilon)f^{(2)}(x) = h'(x)\begin{Vmatrix} (1-\epsilon)I + O(\epsilon) & O(\epsilon) \\ O(\epsilon) & \epsilon M_3(\xi') \end{Vmatrix}^{-1} h(x) \\ &\quad - f^{(2)'}(x)[\epsilon M_3(\xi')]^{-1}f^{(2)}(x) \end{aligned}$$

where in each matrix we have used the fact that $f^{(2)} = 0$ on the support of ξ_s.

In each case $O(\epsilon)$ is a matrix which approaches 0 as $\epsilon \to 0$. Therefore

$$d_{s\mid k}(x, \xi_\epsilon) = (1-\epsilon)^{-1}h^{(1)\prime}(x)h^{(1)}(x) + h'(x)O(\epsilon)h(x).$$

Hence

$$s \leqq \max_x d_{s\mid k}(x, \xi_\epsilon) \leqq (1-\epsilon)^{-1} \max_x h^{(1)\prime}(x)h^{(1)}(x) + \max_x h'(x)O(\epsilon)h(x) \\ \to \max_x h^{(1)\prime}(x)h^{(1)}(x) \leqq s,$$

the last inequality following from (3.10). Therefore ξ_s is optimal for s out of k parameters, proving the theorem.

We now return to the example of special n-tic regression on the simplex. Let $\mathfrak{X}$ be the q-simplex. Let $f(x)$ be the k-vector whose components are the special monomials of degree $\leqq n$, where $n \leqq q + 1$. Let $f^{(1)}(x)$ be the s-vector consisting of those components of degree $\leqq m$, where $m \leqq n$. And let G be the group of permutations of the coordinates of the simplex. It is easily seen that there is a corresponding $\bar{G}$ so that G and $\bar{G}$ satisfy the conditions of Theorem 3.3. Let $d_s(x, \xi, D)$ be as in Theorem 3.3.

Lemma 3.1. *In the model described above, let $n = m + 1$, and let ξ^* be optimal for $\theta^{(1)}$. Then $d_s(x, \xi^*, D)$ is unbounded on any line determined by setting all but two of the x_i constant and holding $\sum_{i=1}^{q+1} x_i = 1$.*

When $s = k$ the analogous statement about $d(x, \xi^*)$ is trivial, but when $s < k$, $f^{(1)}(x) - D'f^{(2)}(x)$ could conceivably be constant on such a line, making $d_s(x, \xi^*, D)$ constant there.

Proof. Suppose $d_s(x, \xi^*, D)$ is bounded on such a line, say the line L determined by setting $x_i = a_i$, $i < q$. We will show that a contradiction follows.

In general if z is an s-vector and B is a positive definite symmetric $s \times s$ matrix with smallest eigenvalue λ_0, then $|z'Bz| \geqq \lambda_0|z|^2$, where $|\ |$ denotes the absolute value of a number and the Euclidean norm of a vector. Thus if $d_s(x, \xi^*, D)$ is bounded on a line, $|f^{(1)}(x) - D'f^{(2)}(x)|^2$ must be bounded on that line. Hence each of the s components of $f^{(1)}(x) - D'f^{(2)}(x)$ is bounded on that line. Since each component of $f^{(1)}(x) - D'f^{(2)}(x)$ is a polynomial, each component must be constant on that line. Let us write the qth component as

$$x_q - \sum_{i_1<\cdots<i_n} c_{i_1,\cdots,i_n} x_{i_1} \cdots x_{i_n}. \tag{3.11}$$

Here we have used the assumption that $n = m + 1$, so $f^{(2)}$ consists only of terms of degree n. Observe that $n = m + 1 \geqq 2$. To avoid notational problems later, we define $c_{j_1,\cdots,j_n} = c_{i_1,\cdots,i_n}$ whenever $(j_1, \cdots, j_n)$ is a rearrangement of $(i_1, \cdots, i_n)$.

If τ is any permutation of $q + 1$ elements, let L_τ be the image of L under that permutation of the coordinates. L_τ is defined by setting $x_i = a_{\tau(i)}$ for those i with $\tau(i) < q$. We will write $h = \tau^{-1}(q)$ and $l = \tau^{-1}(q + 1)$, and write $x_h = u$ and $x_l = v$. This is simply a notational convenience. Thus u and v are variables satisfying $u + v + \sum_1^{q-1} a_i = 1$.

We assumed $d_s(x, \xi^*, D)$ to be bounded on L. By the invariance of $d_s(x, \xi^*, D)$ it is therefore bounded on L_τ for every τ. Therefore on every L_τ, (3.11) is constant.

For any τ, on the line L_τ, (3.11) becomes

$$
\begin{aligned}
(3.12)\qquad x_q - [&\textstyle\sum_{i_1<\cdots<i_n} c_{i_1,\cdots,i_n} a_{\tau(i_1)} \cdots a_{\tau(i_n)} \\
&+ \textstyle\sum_{i_1<\cdots<i_{n-1}} c_{i_1,\cdots i_{n-1},h} a_{\tau(i_1)} \cdots a_{\tau(i_{n-1})} u \\
&+ \textstyle\sum_{i_1<\cdots<i_{n-1}} c_{i_1,\cdots,i_{n-1},l} a_{\tau(i_1)} \cdots a_{\tau(i_{n-1})} v \\
&+ \textstyle\sum_{i_1<\cdots<i_{n-2}} c_{i_1,\cdots,i_{n-2},h,l} a_{\tau(i_1)} \cdots a_{\tau(i_{n-2})} uv]
\end{aligned}
$$

where each summation is taken over $i_j \neq h, l$ and where $x_q = a_{\tau(q)}$, u, or v, depending on whether $\tau(q) < q$, $= q$, or $= q + 1$. Suppose for the moment that $n > 2$. If $n = 2$ much of what follows simplifies, as discussed at the end of the proof.

Our first goal is to prove

$$(3.13)\qquad \textstyle\sum_{i_1<\cdots<i_{n-1}, i_j \neq l} c_{i_1,\cdots,i_{n-1},l} = 0, \quad \textit{for all} \quad l,\ 1 \leqq l \leqq q + 1.$$

Since (3.11) is constant on L_τ, the expression (3.12) is a constant, (the value of which may depend on τ). Therefore the coefficients of the quadratic term uv must total zero:

$$\textstyle\sum_{i_1<\cdots<i_{n-2}, i_j\neq h,l} c_{i_1,\cdots,i_{n-2},h,l} a_{\tau(i_1)} \cdots a_{\tau(i_{n-2})} = 0, \qquad \textit{for} \quad h = \tau^{-1}(q),\ l = \tau^{-1}(q + 1).$$

Let $B_{h,l} = \{\tau \mid h = \tau^{-1}(q),\ l = \tau^{-1}(q + 1)\}$. Summing over all τ in $B_{h,l}$, we get

$$(3.14)\qquad \textstyle\sum_{i_1<\cdots<i_{n-2}, i_j\neq h,l} c_{i_1,\cdots,i_{n-2},h,l} \sum_{\tau\varepsilon B_{h,l}} a_{\tau(i_1)} \cdots a_{\tau(i_{n-2})} = 0.$$

In the sum over $B_{h,l}$, each possible product $a_{j_1} \cdots a_{j_{n-2}}$, with $j_1 < \cdots < j_{n-2} < q$, occurs the same number of times, by symmetry. A direct combinatorial argument therefore gives

$$(3.15)\qquad \textstyle\sum_{i_1<\cdots<i_{n-2}, i_j\neq h,l} c_{i_1,\cdots,i_{n-2},h,l} \sum_{j_1<\cdots<j_{n-2}<q} (q + 1 - n)! \cdot (n - 2)!\, a_{j_1} \cdots a_{j_{n-2}} = 0$$

for any $h \neq l$, $1 \leqq h \leqq q + 1$, $1 \leqq l \leqq q + 1$. We would like to show that the sum involving the a's is nonzero. Since each $a_{j_i} \geqq 0$ (the a_i are coordinates in a simplex), we need only show that some $n - 2$ of the a_{j_i} are positive; then the sum will be positive. If no $n - 2$ of the a's were positive, (3.12) would reduce to x_q. Choosing τ with $\tau(q) = q$ would make $x_q = u$, a variable. But (3.12) is constant for any τ. Therefore some $n - 2$ of the a's are positive, and (3.15) reduces to

$$\textstyle\sum_{i_1<\cdots<i_{n-2}, i_j\neq h,l} c_{i_1,\cdots,i_{n-2},h,l} = 0, \quad \text{for all} \quad h \neq l.$$

We fix l and sum over $h \neq l$ to get

$$(n - 1) \textstyle\sum_{i_1<\cdots<i_{n-1}, i_j\neq l} c_{i_1,\cdots,i_{n-1},l} = 0,$$

which reduces to (3.13), the desired equation. We will use (3.13) shortly.

Now let us return to (3.12). We have examined the quadratic terms; we now examine the linear terms.

It turns out that we only need to consider those τ with $\tau(q) = q$, $\tau(q+1) = q+1$. That is, $h = q$, $l = q+1$. Since (3.12) is constant, the linear terms of (3.12) must total a constant $K(\tau)$. Thus

$$u - [\sum_{i_1<\cdots<i_{n-1}, i_j<q} c_{i_1,\cdots,i_{n-1},q} a_{\tau(i_1)} \cdots a_{\tau(i_{n-1})} u + \sum_{i_1<\cdots<i_{n-1}, i_j<q} c_{i_1,\cdots,i_{n-1},q+1} a_{\tau(i_1)} \cdots a_{\tau(i_{n-1})} v] = K(\tau).$$

Summing over the set C (say) of all τ which leave q and $q+1$ fixed, we get

$$(q-1)!\,u - [\sum c_{i_1,\cdots,i_{n-1},q} \sum_{\tau\varepsilon C} a_{\tau(i_1)} \cdots a_{\tau(i_{n-1})} u + \sum c_{i_1,\cdots,i_{n-1},q+1} \sum_{\tau\varepsilon C} a_{\tau(i_1)} \cdots a_{\tau(i_{n-1})} v] = \sum_{\tau\varepsilon C} K(\tau).$$

Now in the same way as we went from (3.14) to (3.15) we see that for any $i_1, \cdots, i_{n-1}$ each product $a_{j_1} \cdots a_{j_{n-1}}$, with $j_1 < \cdots < j_{n-1}$, occurs in the sum over C exactly $(q-n)!\,(n-1)!$ times. Thus we get

$$(q-1)!\,u - [\sum c_{i_1,\cdots,i_{n-1},q} \sum_{j_1<\cdots<j_{n-1}} (q-n)!\,(n-1)!\,a_{j_1} \cdots a_{j_{n-1}} u + \sum c_{i_1,\cdots,i_{n-1},q+1} \sum_{j_1<\cdots<j_{n-1}} (q-n)!\,(n-1)!\,a_{j_1} \cdots a_{j_{n-1}} v] = \sum_{\tau\varepsilon C} K(\tau).$$

With the obvious definition of A and K, this is

$$(3.16)\quad u - A[\sum_{i_1<\cdots<i_{n-1}, i_j<q} c_{i_1,\cdots,i_{n-1},q} u + \sum_{i_1<\cdots<i_{n-1}, i_j<q} c_{i_1,\cdots,i_{n-1},q+1} v] = K.$$

Now

$$\sum_{i_1<\cdots<i_{n-1}, i_j<q} c_{i_1,\cdots,i_{n-1},q+1} = \sum_{i_1<\cdots<i_{n-1}, i_j\neq q+1} c_{i_1,\cdots,i_{n-1},q+1} - \sum_{i_1<\cdots<i_{n-2}, i_j<q} c_{i_1,\cdots,i_{n-2},q,q+1}.$$

At last we use equation (3.13) here, letting $l = q+1$, to get

$$\sum_{i_1<\cdots<i_{n-1}, i_j<q} c_{i_1,\cdots,i_{n-1},q+1} = -\sum_{i_1<\cdots<i_{n-2}, i_j<q} c_{i_1,\cdots,i_{n-2},q,q+1}.$$

There is a similar identity if we reverse the roles of q and $q+1$. Therefore (3.16) becomes

$$u - A[-\sum_{i_1<\cdots<i_{n-2}, i_j<q} c_{i_1,\cdots,i_{n-2},q,q+1}(u+v)] = K.$$

But $u + v = 1 - \sum a_i$, a constant. Therefore u is a constant, a contradiction.

The proof is much simpler if $n = 2$. In that case the sole quadratic coefficient of (3.12) is $c_{h,l}$, which must be zero because (3.12) is constant. This holds for any $h \neq l$. Therefore (3.12) immediately reduces to x_q. This is constant, but if $\tau(q) = q$, $x_q = u$, a variable. This is the desired contradiction.

Therefore $d_s(x, \xi^*, D)$ cannot be bounded on L.

It is not clear whether the conclusion of the lemma holds for other values of m and n. If $m = 1$, $n = 3$, and $q + 1 = 3$, it is not hard (at least compared to the proof just completed) to show that the conclusion holds. However setting $q + 1 = 3$ is such a great simplification that this result probably gives no clue about the general case.

We obtain a result from the lemma which is given as a theorem.

THEOREM 3.5. *In the model described above, special n-tic regression on the simplex with* $n = m + 1$, *the support of any optimal design for* $\theta^{(1)}$ *is a subset of the barycenters.*

For an optimal design ξ^* we use $d_s(x, \xi^*, D)$ as we used $d(x, \xi^*)$ in the proof of Theorem 2.2. On the relevant lines L, unboundedness of $d_s(x, \xi^*, D)$ follows from the lemma just proved, and symmetry from Theorem 3.3. Thus it is not hard to show that $d_s(x, \xi^*, D)$ can attain its maximum in the simplex only at the barycenters, where the maximum is s. Since $\int d_s(x, \xi^*, D)\, d\xi^* = s$, ξ^* must be supported on a subset of the barycenters. The details are left to the reader.

The result just proved is much weaker than the desired statement which would involve no restriction on n and m. However as has been noted the difficulty lies in obtaining the conclusion of the lemma without assuming such a restriction.

For the rest of this section we will investigate the optimality of ξ^0, introduced in Section 2, when estimating $\theta^{(1)}$ in the case of special n-tic regression on the simplex. Let $\mathfrak{X}$, f and $f^{(1)}$ be as described just before the lemma. Let $z_1, \cdots, z_s$ denote the barycenters of depth $\leqq m$, and $z_{s+1}, \cdots, z_k$ those of depth from $m + 1$ to n. Let L, of the form (3.8), be the nonsingular $k \times k$ matrix defined explicitly by letting the ith component of $g(x) = Lf(x)$ be (2.1). Then $g_i(z_j) = \delta_{ij}$, $1 \leqq i, j \leqq k$. Let ξ^0 be the design uniform on $\{z_1, \cdots, z_s\}$.

THEOREM 3.6. *If* $m \leqq 3$, ξ^0 *is optimal for* $\theta^{(1)}$.

PROOF. This is immediate from Theorem 3.4 (setting $A = I$ and $B = 0$ there) and the facts stated just before Theorem 2.4.

THEOREM 3.7. *If* $n = q + 1$, ξ^0 *is optimal for* $\theta^{(1)}$.

PROOF. Let ξ' be uniform on $\{z_{s+1}, \cdots, z_k\}$, and let $\xi_\epsilon = (1 - \epsilon)\xi^0 + \epsilon\xi'$. Then

$$\begin{aligned} d_{s\,|\,k}(x, \xi_\epsilon) &= g'(x)[\textstyle\int gg'\, d\xi_\epsilon]^{-1}g(x) - g^{(2)\prime}(x)[\textstyle\int g^{(2)}g^{(2)\prime}\, d\xi_\epsilon]^{-1}g^{(2)}(x) \\ &= g'(x)\begin{Vmatrix} (1-\epsilon)s^{-1}I & 0 \\ 0 & \epsilon r^{-1}I \end{Vmatrix}^{-1} g(x) - g^{(2)\prime}(x)[\epsilon r^{-1}I]^{-1}g^{(2)}(x), \end{aligned}$$

where $(1 - \epsilon)s^{-1}I$ is $s \times s$ and $\epsilon r^{-1}I$ is $r \times r$. Thus

$$d_{s|k}(x, \xi_\epsilon) = s(1 - \epsilon)^{-1} \textstyle\sum_{i=1}^{s} g_i^2(x).$$

This is symmetric, nonnegative and at most quadratic in each variable x_i. Therefore it attains its maximum at a barycenter. (If the expression is constant on some line, it may also attain its maximum elsewhere.) But since $n = q + 1$, the only barycenters are $z_1, \cdots, z_k$, and there $\sum g_i^2(x) = 0$ or 1. Therefore $\lim_\epsilon \max_x d_{s|k}(x, \xi_\epsilon) = s$, and ξ^0 is optimal by Theorem 3.1.

The last theorem about ξ^0 gives a case in which it is not optimal.

THEOREM 3.8. *For* $q + 1 - n$ *fixed* >0 *and* $n - m$ *fixed* >0 *and* n *sufficiently large,* ξ^0 *is not optimal for* $\theta^{(1)}$.

If we only wished to prove the theorem we would merely have to find some design better than ξ^0, and the proof below could be simplified. Instead we consider a more complicated design, which in the special case $m + 1 = n = q$ is the most general symmetric competitor of ξ^0. Thus we not only prove the theorem, but also determine in this special case how large n must be before ξ^0 is not optimal.

PROOF. ξ^0 is uniform on $\{z_1, \cdots, z_s\}$. Let ξ^+ be any symmetric design on $\{z_1, \cdots, z_s\}$, ξ' uniform on $\{z_{s+1}, \cdots, z_k\}$, and ξ_0 concentrated on z_{k+1}, a barycenter of depth $n+1$. Define $\xi = \xi(b, c) = a\xi^+ + b\xi' + c\xi_0$, where $a = 1 - b - c$ and $a, b, c \geqq 0$. (If $m + 1 = n = q$, $\xi(b, c)$ will be the most general symmetric competitor of ξ^0.) Without loss of generality we may consider the regression functions to be the g_i defined above. Then

$$M([1-\alpha]\xi^0 + \alpha\xi) = (1-\alpha)s^{-1}\begin{Vmatrix} I & 0 \\ 0 & 0 \end{Vmatrix} + \alpha a \begin{Vmatrix} M_1(\xi^+) & 0 \\ 0 & 0 \end{Vmatrix} + \alpha b r^{-1}\begin{Vmatrix} 0 & 0 \\ 0 & I \end{Vmatrix} + \alpha c g(z_{k+1})g'(z_{k+1}).$$

There exist orthogonal matrices R_1 and R_2 with $(R_1 g^{(1)}(z_{k+1}))' = (l_1, 0, \cdots, 0)$ and $(R_2 g^{(2)}(z_{k+1}))' = (l_2, 0, \cdots, 0)$ where

$$l_1^2 = \textstyle\sum_{i=1}^{s} g_i^2(z_{k+1}) = \sum_{p=1}^{m} \binom{n+1}{p}(p/(n+1))^{2n} \tag{3.17}$$

and

$$l_2^2 = \textstyle\sum_{i=s+1}^{k} g_i^2(z_{k+1}) = \sum_{p=m+1}^{n} \binom{n+1}{p}(p/(n+1))^{2n}. \tag{3.18}$$

(The evaluation of the sums is as in the proof of Theorem 2.4.) Write

$$R = \begin{Vmatrix} R_1 & 0 \\ 0 & R_2 \end{Vmatrix}.$$

Suppressing the argument of $M([1-\alpha]\xi^0 + \alpha\xi)$, we have $\det M^* = \det (RMR')^*$. After straightforward manipulation, $(RMR')^*$ reduces to

$$(1-\alpha)s^{-1}I + \alpha\left[aR_1M_1(\xi^+)R_1' + \begin{Vmatrix} d & 0 \\ 0 & 0 \end{Vmatrix}\right]$$

where $d = bcl_1^2/(b + rcl_2^2)$. Let P be generic for a polynomial. We have

$$\det M^* = (1-\alpha)^s s^{-s} + (1-\alpha)^{s-1}s^{-s+1}\alpha[a \operatorname{tr} R_1M_1(\xi^+)R_1' + d] + \alpha^2 P(\alpha).$$

The trace of $R_1M_1(\xi^+)R_1'$ equals 1, because R_1 is orthogonal and $g_i(z_j) = \delta_{ij}$ for $1 \leqq i, j \leqq s$. Hence

$$\det M^* = (1-\alpha)^{s-1}s^{-s}[1 - \alpha + \alpha s a + \alpha s d + \alpha^2 P(\alpha)].$$

Thus

$$\frac{\partial}{\partial\alpha}\log\det M^* \mid_{\alpha=0^+} = s(-1 + a + d) = s(bcl_1^2/(b + rcl_2^2) - b - c).$$

As shown by Kiefer (1961, page 304 ff.), this derivative is $\leqq 0$ if ξ^0 is optimal, and only if ξ^0 is at least as good as $\xi(b, c)$. Now the derivative is $\leqq 0$ for all b,

$c \geqq 0$ with $b + c \leqq 1$ if and only if for all those b and c

$$0 \leqq b^2 - (l_1^2 - 1 - rl_2^2)bc + rl_2^2c^2,$$

which holds if and only if

$$l_1^2 - (1 + r^{\frac{1}{2}}l_2)^2 \leqq 0. \tag{3.19}$$

We now show that for $q + 1 - n$ fixed >0 and $n - m$ fixed >0, $l_1^2 - (1 + r^{\frac{1}{2}}l_2)^2 \to \infty$ as $n \to \infty$, and therefore for n sufficiently large ξ^0 is not optimal. Write $q - n = t \geqq 0, n - m = u \geqq 1$.

Write $\binom{n+1}{p}(p/(n+1))^{2n} = G_p$. Then $l_2^2 = \sum_{p=n-u+1}^{n} G_p$. To bound l_2^2 note that for n large, G_{n-j}/G_{n-j+1} is asymptotic to $ne^{-2}/(j+1) \geqq 1$. So $l_2^2 \leqq uG_{n-u+1}$ for n large. Now

$$r = \sum_{p=m+1}^{n} \binom{q+1}{p} \leqq u\binom{q+1}{m+1} \quad \text{for} \quad n \quad \text{large}$$

which equals $u\binom{n+t+1}{n-u+1}$.

Also, $l_1^2 = \sum_{p=1}^{n-u} G_p$. Let G_{n-v} be a typical term. Then

$$\frac{l_1^2}{rl_2^2} \geqq G_{n-v}/u\binom{n+t+1}{n-u+1}uG_{n-u+1} = \frac{(n-u+1)!(n-u+1)!}{(n-v)!(n+t+1)!} \cdot \frac{(t+u)!u!}{(v+1)!u^2} \cdot \left(\frac{n-v}{n-u+1}\right)^{2n}$$

which we break up into three parts as shown. If we fix v sufficiently large ($v \geqq 2u + t$ is enough), and let $n \to \infty$, application of Stirling's formula shows that the first ratio tends to infinity. The second is constant and the third approaches a fixed power of e. Therefore $l_1^2/rl_2^2 \to \infty$. Also $l_1^2 \to \infty$, so $l_1^2 - (1 + r^{\frac{1}{2}}l_2)^2 \to \infty$, which proves the assertion.

In the case $m + 1 = n = q$, the most general symmetric competitor of ξ^0 is a design of the form $\xi(b, c)$. There is certainly some symmetric optimal design, so in this case ξ^0 is optimal if and only if it is at least as good as every $\xi(b, c)$, hence if and only if (3.19) holds. A computer calculation of the expression in (3.19) gives the result that ξ^0 is optimal for $m + 1 = n = q \leqq 50$, and not optimal for $m + 1 = n = q = 51$ through 55. Further calculations were not performed.

If $m > 3$ and if $n - m > 1$ or $q > n$ it is not known whether ξ^0 is ever optimal, although it seems likely to be optimal for $n - m$ and $q - n$ and n all small. It is also not known how large n must be before a design of the form $\xi(b, c)$ is better than ξ^0. However ξ^0 is better than $\xi(b, c)$ if (3.19) holds, so n must certainly be greater than 50. For (see (3.17) and (3.18))

$$l_1^2 - (1 + r^{\frac{1}{2}}l_2)^2 = \sum_{p=1}^{m} G_p - [1 + r^{\frac{1}{2}}(\sum_{p=m+1}^{n} G_p)^{\frac{1}{2}}]^2. \tag{3.20}$$

But $r = \sum_{p=m+1}^{n} \binom{q+1}{p} \geqq \sum_{p=m+1}^{n} \binom{n+1}{p} \geqq n + 1$, so the expression on the right in (3.20) is no greater than

$$\sum_{p=1}^{n-1} G_p - [1 + (n+1)^{\frac{1}{2}}G_n^{\frac{1}{2}}]^2.$$

This last expression is $l_1^2 - (1 + r^{\frac{1}{2}}l_2)^2$ for the case $m + 1 = n = q$, and this is negative for $n \leqq 50$.

4. Efficiency of various designs. In Section 1 upper bounds were given on the number of points needed for estimating θ or $\theta^{(1)}$. In this section we give examples in which the bounds are attained. In each model we also show that there are always designs supported on a small number of points (k or fewer) with D-efficiency bounded below by quantities which will be given. Unfortunately these bounds are sharp only in a few cases. However we do give a sharp lower bound on the G-efficiency of the best k-point design for estimating $\theta' f(x)$. If ξ is a design for estimating θ, we obtain bounds in both directions relating the D-efficiency and G-efficiency of ξ, including the result that any ξ has D-efficiency no less than its G-efficiency. These last relations are slight improvements over known results; however we also obtain corresponding relations if ξ is a design for estimating $\theta^{(1)}$. Finally there is a discussion of the D-efficiency and G-efficiency of a single design used in several models, and an application of this to product designs considered by Hoel.

We begin with two examples mentioned above. First is an example in which $n_0 = k(k + 1)/2$ points are needed for an optimal design for θ. This example is similar to the example which is concerned with optimality for $\theta^{(1)}$, but much simpler. The second example does not reduce to the first when we set $s = k$, and the verification of the properties of the second example uses the fact that $s < k$. Thus the examples are similar but distinct.

EXAMPLE 4.1. If $k = 1$ the question is trivial. For $k > 1$, let

$$\mathfrak{X} = \{x_{ij} \mid i \leqq j \leqq k\},$$

so $\mathfrak{X}$ has n_0 points. Let α be chosen with $2(2k - 4)/(2k - 3) < \alpha < 2$. The reason for this choice will become apparent. Let $f_i(x_{jj}) = \alpha^{\frac{1}{2}}\delta_{ij}$, and $f_i(x_{lj}) = \delta_{il} + \delta_{ij}$ for $l \neq j$. We now verify that the only optimal design is supported on all n_0 points of $\mathfrak{X}$.

Writing ξ_{ij} for $\xi(x_{ij})$ we have for any ξ

$$M_{ii}(\xi) = \xi_{ii}\alpha + \textstyle\sum_{l<i} \xi_{li} + \sum_{i<j} \xi_{ij},$$
$$M_{ij}(\xi) = M_{ji}(\xi) = \xi_{ij} \quad \text{for} \quad i < j.$$

Since $M(\xi)$ is the same for all optimal ξ, it is immediate that the optimal ξ is unique. It is therefore invariant, and we need henceforth consider only designs which are invariant in the following sense.

For any permutation π on k elements define $gx_{ij} = x_{\pi_i \pi_j}$ or $x_{\pi_j \pi_i}$. (Just one of these will be in $\mathfrak{X}$, as $\pi_i \leqq \pi_j$ or $\pi_i \geqq \pi_j$.) Define $\bar{g}\theta_i = \theta_{\pi_i}$. Then $(\bar{g}\theta)' f(gx) = \theta' f(x)$, so the problem is invariant under such g. The invariant designs are those of the form $\xi_{ii} = a$ and $\xi_{ij} = b$ for $i < j$. Note $ka + k(k - 1)b/2 = 1$.

For such designs $M(\xi) = (a\alpha + (k - 2)b)I + bU$, where U is a $k \times k$ matrix with every entry 1. Then $\det M(\xi) = [a\alpha + 2(k - 1)b][a\alpha + (k - 2)b]^{k-1}$. After substituting for b by the relation $ka + k(k - 1)b/2 = 1$, the derivative of this determinant with respect to a at $a = 0$ is positive if $\alpha > 2(2k - 4)/(2k - 3)$, and the derivative with respect to a at $b = 0$ is negative if $\alpha < 2$. Thus for optimality we must have $a \neq 0$ and $b \neq 0$, and all n_0 points are needed for an optimal design.

In the second example $n_0 = s(s+1)/2 + rs$ points are needed for an optimal design for $\theta^{(1)}$.

EXAMPLE 4.2. Let $\mathfrak{X} = \{x_{ij} \mid i \leqq j \leqq s\} \cup \{y_{ij} \mid i \leqq s < j \leqq k\}$. Assume $s < k$. There are $n_0 = s(s+1)/2 + rs$ points *in* $\mathfrak{X}$. Choose γ so that

$$0 < \gamma \quad \text{if} \quad s = 1,$$

$$0 < \gamma < 1/(3r) \quad \text{if} \quad s = 2;$$

$0 < \gamma$ and γ small enough so that

$$(1/(r\gamma) - 1)^2 > 2(s-1)^2/(s-2) \quad \text{if} \quad s > 2.$$

Now let

$$f_i(x_{jj}) = \delta_{ij} \quad \text{if} \quad i \leqq s$$

$$f_i(x_{jj}) = \gamma \quad \text{if} \quad i > s$$

$$f_i(x_{lj}) = \delta_{il} + \delta_{ij} \quad \text{if} \quad l \neq j$$

$$f_i(y_{lj}) = \delta_{il} + \delta_{ij} \quad \text{if} \quad l \leqq s < j.$$

The reader is referred to Atwood (1968) for the proof. Here is the idea. If $s = 1$ it is easy to show that all n_0 points are needed to estimate $\theta^{(1)}$, hence they are necessary for optimality. For $s > 1$, it is not too hard to show that any invariant optimal design requires all n_0 points of support; the method is the same as was used for Example 4.1. Finally a long series of computations, treating different cases separately, shows that any optimal design must be invariant.

We now consider the efficiency of designs with fewer points of support.

When estimating θ, we define the *D-efficiency* of a design ξ as

$$[\det M(\xi)/\max_{\xi'} \det M(\xi')]^{1/k}.$$

To justify this definition, suppose we take m observations using ξ and n observations using an optimal design ξ^*. The best linear unbiased estimators thus obtained have covariance matrices $\sigma^2 m^{-1} M^{-1}(\xi)$ and $\sigma^2 n^{-1} M^{-1}(\xi^*)$ respectively. These matrices will have equal determinants if and only if the ratio of sample sizes $n/m = [\det M(\xi)/\det M(\xi^*)]^{1/k}$.

Likewise we define the G-efficiency of ξ to be $k/\max_x d(x, \xi)$. And when estimating $\theta^{(1)}$ we define the *D-efficiency* of ξ to be $[\det M^*(\xi)/\max_{\xi'} \det M^*(\xi')]^{1/s}$. These definitions are also justified as ratios of sample sizes.

The first theorem gives a lower bound on the D-efficiency of certain designs supported on k or fewer points. D. Meeter (1967) has pointed out that this result was obtained earlier in the case $s = k$ by M. J. Box (1968). It was found independently by the author, and the case $s < k$ given here is not considered by Box or Meeter.

THEOREM 4.1. *For $s \leqq k$, let η be a D-optimal design for $\theta^{(1)}$, supported on n points where $n \leqq n_0 = s(s+1)/2 + rs$. Let $M(\eta)$ have rank m, $s \leqq m \leqq k$. There is a design ζ uniform on m points of the support of η satisfying*

$$\det M^*(\zeta)/\det M^*(\eta) \geqq n^s/[\tbinom{n}{m}\tbinom{m}{s}m^s] \geqq n_0^s/[\tbinom{n_0}{m}\tbinom{m}{s}m^s].$$

If $s > 1$ *this in turn is bounded below by* $n_0^s/[\binom{n_0}{k}\binom{k}{s}k^s]$. *If* $s = 1$ *it is bounded below by* $n_0/[\binom{n_0}{m_0}m_0^2]$, *where* m_0 *is the smallest positive integer satisfying*

$$m_0^2 \geqq (k - m_0)(m_0 + 1).$$

We shall comment extensively on the sharpness of these bounds, and on the asymptotic values of the bounds as the various parameters get large. One such result is given here as a corollary.

COROLLARY. *When* $s > 1$, *the best design for* $\theta^{(1)}$ *on* k *or fewer points has D-efficiency at least* $n_0k^{-1}[\binom{n_0}{k}\binom{k}{s}]^{-1/s}$. *If* r *is fixed and* $s \to \infty$, *this bound has limit* e^{-1}.

PROOF OF THEOREM. First we remark that if we operate on any $M(\xi)$ which has nonsingular $M_3(\xi)$ by a matrix

$$L = \begin{Vmatrix} I & -M_2M_3^{-1} \\ 0 & I \end{Vmatrix}$$

we get immediately $\det M(\xi) = \det M^*(\xi) \det M_3(\xi)$.

The problem will be easier if we reformulate it in terms of matrices of full rank. Because $m = \operatorname{rank} M(\eta)$, the $r \times n$ matrix having columns $f^{(2)}(x)$, with x in the support of η, has rank t, with $t \geqq m - s$. We can multiply f by a matrix of the form (3.8), in fact with $L_1 = I$ and $L_2 = 0$, to get a vector Lf with $(Lf)_i(x) = 0$ for $i > t + s$ and x in the support of η. Let g be the $(s + t)$-dimensional column vector with components $(Lf)_1, \cdots, (Lf)_{s+t}$, and let $g^{(1)}$ be the vector consisting of the first s components, $g^{(2)}$ the remaining t components. (If $t = 0$, the following simplifies.) Now for ξ supported on a subset of the support of η, $[\int gg'\, d\xi]^* = [LM(\xi)L']^* = L_1M^*(\xi)L_1' = M^*(\xi)$. In particular if $\xi = \eta$, then by the remark at the beginning of the proof

$$\det \int gg'\, d\eta = \det M^*(\eta) \det \int g^{(2)}g^{(2)\prime}\, d\eta.$$

The two matrices on the right are of full rank, while that on the left is of full rank only if $s + t = m$. Therefore $s + t = m$.

Throughout the rest of the proof we will use the following notation. Let the support of η be written $\{x_1, \cdots, x_n\}$. If M is an m-element subset of $\{1, \cdots, n\}$, then η_M and ζ_M will denote the m-fold products $\prod \eta(x_i)$ and $\prod \zeta(x_i)$ for $i \varepsilon M$, and $\|g(x_M)\|$ will denote the $m \times m$ matrix with entries $g_i(x_j)$, for g_i a component of g and $j \varepsilon M$. If S and T are subsets of $\{1, \cdots, n\}$ with s and t elements respectively, similar definitions will hold for η_S, ζ_S, η_T, ζ_T, $\|g^{(1)}(x_S)\|$, and $\|g^{(2)}(x_T)\|$.

Consider now the case $m = s$. In this case g and $g^{(1)}$ coincide, and $M^*(\eta) = \int gg'\, d\eta$. By the Theorem of Corresponding Minors (see Householder (1964, page 14)),

$$\det M^*(\eta) = \textstyle\sum_M \eta_M \det^2 \|g(x_M)\|.$$

Let M_0 be that M which maximizes $\det^2 \|g(x_M)\|$, and let ζ be uniform on $\{x_i \mid i \varepsilon M_0\}$. Then

$$\det M^*(\zeta) = m^{-m} \max_M \det^2 \|g(x_M)\|.$$

Therefore

$$(4.1)\quad \det M^*(\zeta)/\det M^*(\eta) = m^{-m} \max_M \det^2 \|g(x_M)\|/\sum_M \eta_M \det^2 \|g(x_M)\| \geqq m^{-m}/\sum_M \eta_M .$$

It is not hard to show (see Keilson (1966)) that if $\phi(x_1, \cdots, x_n)$ is a symmetric multilinear function on the $(n-1)$-simplex, then ϕ attains its extrema among the barycenters. So, putting $\phi(\eta_1, \cdots, \eta_n) = \sum_M \eta_M$, we get

$$(4.2)\quad \det M^*(\zeta)/\det M^*(\eta) \geqq m^{-m}/\max_j \binom{j}{m} j^{-m}, \qquad \textit{where} \quad m \leqq j \leqq n.$$

Now we have the ratio

$$\binom{j}{m} j^{-m}/\binom{j-1}{m}(j-1)^{-m} = (1 - 1/j)^m/(1 - m/j) \geqq 1,$$

the last inequality being well known. Therefore in (4.2) the maximum occurs when j is as large as possible. This proves that when $m = s$, we have

$$\det M^*(\zeta)/\det M^*(\eta) \geqq n^m/[\binom{n}{m} m^m] \geqq n_0^{\,m}/[\binom{n_0}{m} m^m].$$

Let us now consider the case $m = s + t$, $t > 0$. In this case we have

$$\det M^*(\eta) = \det \int gg'\, d\eta/\det \int g^{(2)} g^{(2)\prime}\, d\eta = \sum_M \eta_M \det^2 \|g(x_M)\|/\sum_T \eta_T \det^2 \|g^{(2)}(x_T)\|.$$

In the numerator we will delete any zero terms. Then

$$(4.3)\quad \det M^*(\eta) \leqq \sum_M [\eta_M \det^2 \|g(x_M)\|/\sum_{T\subset M} \eta_T \det^2 \|g^{(2)}(x_T)\|],$$

where from now on we index only over those M for which the numerator of the summand is nonzero. Because each numerator is nonzero, consideration of the expansion of $\det \|g(x_M)\|$ shows that each denominator is nonzero. The right hand expression in (4.3) equals

$$(4.4)\quad \begin{aligned} &\sum_M [\det^2 \|g(x_M)\|/\sum_{T\subset M, S=M-T} (\eta_S)^{-1} \det^2 \|g^{(2)}(x_T)\|] \\ &\leqq \sum_M [\det^2 \|g(x_M)\|/\min_{S\subset M} (\eta_S)^{-1} \sum_{T\subset M} \det^2 \|g^{(2)}(x_T)\|] \\ &= \sum_M \max_{S\subset M} \eta_S [\det^2 \|g(x_M)\|/\sum_{T\subset M} \det^2 \|g^{(2)}(x_T)\|]. \end{aligned}$$

Let M_0 be that M which maximizes the expression in square brackets in the last line of (4.4). Let ζ be uniform on $\{x_i \mid i \in M_0\}$. Then

$$\det M^*(\zeta) = m^{-s} \max_M [\det^2 \|g(x_M)\|/\sum_{T\subset M} \det^2 \|g^{(2)}(x_T)\|].$$

Therefore

$$(4.5)\quad \det M^*(\zeta)/\det M^*(\eta) \geqq m^{-s}/\sum_M \max_{S\subset M} \eta_S .$$

We have been summing over M as restricted just below (4.3). We now again sum over all m-element sets M; this can only decrease the right hand side of (4.5). Since for any M

$$(4.6)\quad \max_{S\subset M} \eta_S \leqq \sum_{S\subset M} \eta_S ,$$

we have

$$(4.7)\qquad \det M^*(\zeta)/\det M^*(\eta) \geqq m^{-s}/\textstyle\sum_M \sum_{S\subset M} \eta_S .$$

The denominator on the right is a constant (depending on n) times $\sum_S \eta_S$, which was shown in the proof for the case $m = s$ to be maximized when η is uniform on as large a set as possible. Therefore

$$\det M^*(\zeta)/\det M^*(\eta) \geqq n^s/[\tbinom{n}{m}\tbinom{m}{s}m^s] \geqq n_0^s/[\tbinom{n_0}{m}\tbinom{m}{s}m^s].$$

We now only need to prove that for $s \leqq m \leqq k$,

$$B(m) = n_0^s/[\tbinom{n_0}{m}\tbinom{m}{s}m^s]$$

is bounded below by the quantities given in the statement of the theorem. The ratio $B(m)/B(m-1)$ is increasing, and $B(k)/B(k-1)$ is less than 1 if $s \geqq 2$. Thus if $s \geqq 2$, $B(m)$ is minimized by $m = k$. If $s = 1$, $B(m+1)/B(m) = m^2/(k-m)(m+1)$. This is increasing in m for $m > 0$, and first becomes greater than or equal to 1 when $m = m_0$, defined in the statement of the theorem.

The proof of the corollary is immediate, using Stirling's approximation of the factorial.

In the special case $s = k$, the corollary says that the lower bound on efficiency approaches e^{-1} as $k \to \infty$.

In similar manner it is easy to show that if s is fixed and $r \to \infty$, or if the ratio r/s is fixed and $r \to \infty$, then the lower bound on efficiency approaches 0.

Consideration of the sharpness of the bounds of the theorem is now in order. If $m > s$ strict inequality holds in (4.6), because $\eta(x_i) > 0$, $i = 1, \cdots, n$. Therefore the bound $n^s/[\binom{n}{m}\binom{m}{s}m^s]$ is not attainable, and hence the (smaller) bounds which do not depend on m and n are not attainable.

If $m = s$, the bounds are attainable only if equality holds in (4.1). Equality holds there if and only if $\det^2 \|g(x_M)\|$ is the same for all M. But this can happen only if $m = n - 1$ (or in the trivial cases $m = 0, 1$ or n). To see this, renumber if necessary so that $\|g(x_{M_1})\|$ is nonsingular, where $M_1 = \{1, \cdots, m\}$. Operations on the rows of the $m \times n$ matrix with entries $g_i(x_j)$, (which do not change the ratio $\det^2 \|g(x_M)\|/\det^2 \|g(x_{M'})\|$ for any M or M'), transform $\|g(x_{M_1})\|$ into the $m \times m$ identity matrix, with determinant 1. If $\det^2 \|g(x_M)\|$ is now to equal 1 for every M involving m of the first $m + 1$ columns, $g(x_{m+1})$ must have every entry equal to ± 1. By the same reasoning so must $g(x_{m+2})$. But then any determinant involving columns $m+1$, $m+2$, and all but two of the first m columns is not equal to ± 1.

So when $m = s$ the bound $n^s/[\binom{n}{m}\binom{m}{s}m^s] = n^s/[\binom{n}{s}s^s]$ is attainable only if $m = n - 1$. Since this bound is strictly decreasing in n, the bound $n_0^s/[\binom{n_0}{s}s^s]$ is attainable only if $m = n_0 - 1$, i.e., $s = rs + s(s+1)/2 - 1$. This occurs only if $s \leqq k = 2$. We consider the two possibilities:

If $m = s = 1$, $k = 2$, we have $n_0^s/[\binom{n_0}{s}s^s] = 1$, which is certainly attainable. But since $m_0 = 2$, the lower bound $n_0/[\binom{n_0}{m_0}m_0^2]$ equals $\frac{1}{2} < 1$. Thus the lower bound $n_0/[\binom{n_0}{m_0}m_0^2]$ is not attainable.

If $m = s = k = 2$, the lower bound $n_0^s/[\binom{n_0}{k}k^s]$ is attained. As an example let $\mathfrak{X} = \{x_1, x_2, x_3\}$, let $f_i(x_j) = \delta_{ij}$ for $1 \leqq i, j \leqq 2$, and let $f_i(x_3) = 1$, $i = 1, 2$. It is easy to check that η is uniform on $\mathfrak{X}$ and the bound is attained.

In most situations we will not know n, and if $s < k$ we will generally not know m either. In these cases we can only use the lower bounds which do not depend on these values. Moreover if m is unknown we do not know the number of points of support of ζ, only that the number is between s and k. We might therefore try to find a theorem similar to 4.1 but involving a design supported on s points, or on l points, where l is the number of points necessary to estimate $\theta^{(1)}$. Neither of these ideas works, because of the following facts.

(1) $\theta^{(1)}$ may not be estimable using only s points.

(2) The best l-point design for $\theta^{(1)}$ may be arbitrarily bad, where l is as just defined above.

Both statements are easily proved. There are well known examples to demonstrate (1). One such has been given, Example 4.2 when $s = 1$.

To show (2), let $\mathfrak{X} = \{x_i \mid i = 1, \cdots, k\} \cup \{y_i \mid i = 1, \cdots, s\}$. Let $f_i(x_j) = \delta_{ij}$ for $i \leqq s$ and $j \leqq k$, let $f_i(x_j) = 1$ for $i > s$ and $j \leqq s$, let $f_i(x_j) = \delta_{ij}$ for $i > s$ and $j > s$, and let $f_i(y_j) = \delta_{ij}a$ for $i \leqq k$ and $j \leqq s$, where $|a|$ is small. Then $\theta^{(1)}$ is estimating using s points, $y_1, \cdots, y_s$. As was shown at the end of Section 1, if $\theta^{(1)}$ is estimable using an s-point design then $f^{(2)}$ is zero on the support on such a design, and the best such design is uniform on those s points. Therefore any s-point design which estimates $\theta^{(1)}$ is supported on $\{y_1, \cdots, y_s\}$, and the best such is uniform there. Denoting this design by ξ_s, an optimal design by ξ^*, and the design uniform on $\{x_1, \cdots, x_k\}$ by ξ_k, we have

$$\begin{aligned}\det M^*(\xi_s)/\det M^*(\xi^*) &\leqq \det M^*(\xi_s)/\det M^*(\xi_k)\\ &= a^{2s}s^{-s}/(k^{-s}(1 + rs)^{-1})\end{aligned}$$

which can be arbitrarily small.

We now turn to G-efficiency and give a bound for certain k-point designs.

THEOREM 4.2. *The best* (G) *k-point design* ξ_G *and the best* (D) *k-point design* ξ_D *satisfy*

$$\max_x d(x, \xi_G) \leqq \max_x d(x, \xi_D) \leqq k^2.$$

We could rephrase this to say that the G-efficiency of ξ_G is no less than that of ξ_D, which is no less than $1/k$.

PROOF. We must show the right hand inequality. Let $\{x_1, \cdots, x_k\}$ be the support of ξ_D. Let L be the linear transformation so that $g = Lf$ satisfies $g_i(x_j) = \delta_{ij}$. Then $|g_i(x)| \leqq 1$ for all i and all $x \in \mathfrak{X}$; for if $|g_i(y)| > 1$ then substitution of y for x_i would yield a k-point design ξ' with $\det M(\xi') > \det M(\xi_D)$. Thus we conclude

$$\max_x d(x, \xi_G) \leqq \max_x d(x, \xi_D) = k \max_x \sum g_i^2(x) \leqq k^2.$$

Here is an example in which the bound is attained. Let $\mathfrak{X} = \{x_1, \cdots, x_k, y\}$, let $f_i(x_j) = \delta_{ij}$, and let $f_i(y) = 1$ for $i = 1, \cdots, k$. If ξ is supported on

$\{x_1, \cdots, x_k\}$, $d(x_i, \xi) = 1/\xi_i$ for any i, and $d(y, \xi) = \sum (1/\xi_i) \geqq d(x_i, \xi)$ for any i. So $d(x, \xi)$ is maximized at y. This is minimized by making ξ uniform on $\{x_1, \cdots, x_k\}$. In this case $d(y, \xi) = k^2$. If on the other hand ξ is supported on y and all the x_i's but one, say x_l, let L be the $k \times k$ matrix with entries $a_{ii} = 1$ for all i, $a_{il} = -1$ for $i \neq l$, and $a_{ij} = 0$ otherwise. If $g = Lf$ then $g_i(x_j) = \delta_{ij}$ for $j \neq l$, $g_i(y) = \delta_{il}$, $g_l(x_l) = 1$ and $g_i(x_l) = -1$ for $i \neq l$. Then writing $\xi(y) = \xi_l$, we have $d(x, \xi) = \sum_i g_i^2(x)(1/\xi_i)$. Therefore $d(x_i, \xi) = 1/\xi_i$ if $i \neq l$, $d(y, \xi) = 1/\xi_l$, and $d(x_l, \xi) = \sum (1/\xi_i)$. As before, minimizing over ξ supported on all points but x_l, we have

$$\min_\xi \max_x d(x, \xi) = k^2.$$

This example shows that it is possible to cut the G-efficiency by a factor of k by deleting even a single point from an optimal design. It would therefore seem undesirable in general to simplify a problem by considering only k-point designs, although this has been done in the past, for example by Box and Lucas (1959, page 80).

Since D-efficiency is usually impossible to compute when no optimal design is known, it is of practical value to have relations between the D-efficiency and G-efficiency of a design.

We now develop such relations. If $s = k$, the theorem is a slight improvement over results obtained by Kiefer (1960, p. 389). In any case the method is essentially the same.

Write $\bar{d}(\xi) = \max_x d(x, \xi)$, and $\bar{d}_{s|k}(\xi) = \max_x d_{s|k}(x, \xi)$.

THEOREM 4.3. *Suppose $s \leqq k$, and $M(\xi)$ is nonsingular. Then*

$$[\det M^*(\xi)/\max_{\xi'} \det M^*(\xi')]^{1/s} \geqq s/\bar{d}_{s|k}(\xi).$$

If $\bar{d}(\xi) \leqq c$, where $c > \frac{3}{2}$, then

$$\det M^*(\xi)/\max_{\xi'} \det M^*(\xi') \leqq \exp[-(\bar{d}_{s|k}(\xi) - s)^2/2D]$$

where

$$D = \max\{k - 1 + (c-1)^2, 4(c-1)^2((k-1)/(2c-3)^2 + \tfrac{1}{9}), 2(c-1)(c-s)\}.$$

If $s = k$, the first assertion says that any design ξ has D-efficiency no less than its G-efficiency.

In the second assertion, we might want to let $c = \bar{d}(\xi)$ if this value is known and $> \frac{3}{2}$. Of course if $k > 1$ then $c \geqq k > \frac{3}{2}$ automatically.

PROOF. Let L be of the form (3.8) and such that $LM(\xi)L' = I$. If we let the vector of regression functions be Lf instead of f, then for any ξ', $d_{s|k}(x, \xi')$ and $d(x, \xi')$ are unchanged and $\det M^*(\xi')$ is only multiplied by a constant, $\det^2 L_1$. Thus we may simply consider $M(\xi) = I$ throughout the proof, and will do so.

To prove the first statement of the theorem, let η be D-optimal for $\theta^{(1)}$. Then, using the fact that $M(\xi) = I$,

$$\begin{aligned}\bar{d}_{s|k}(\xi) &\geqq \int d_{s|k}(x, \xi)\, d\eta \\ &= \operatorname{tr} M(\eta) - \operatorname{tr} M_3(\eta) \\ &= \operatorname{tr} M_1(\eta) \\ &\geqq \operatorname{tr} M^*(\eta).\end{aligned}$$

The last inequality is true because $M^*(\eta) = M_1(\eta) - M_2(\eta)M_3^{-1}(\eta)M_2'(\eta)$, (using a limiting value if $M(\eta)$ is singular). Thus, by the arithmetic-geometric mean inequality,

$$\begin{aligned}[\det M^*(\xi)/\det M^*(\eta)]^{1/s} &= [\det M^*(\eta)]^{-1/s} \\ &\geqq s/\operatorname{tr} M^*(\eta) \\ &\geqq s/\bar{d}_{s|k}(\xi).\end{aligned}$$

To prove the second assertion of the theorem, let $\bar{d}_{s|k}(\xi) = s + \epsilon$. Then there is a design ζ with $\int d_{s|k}(x, \xi)\, d\zeta = s + \epsilon$. As in the above paragraph therefore $s + \epsilon = \int d_{s|k}(x, \xi)\, d\zeta = \operatorname{tr} M_1(\zeta)$.

Define $q(\alpha) = \log\det M^*([1 - \alpha]\xi + \alpha\zeta) = \log\det M([1 - \alpha]\xi + \alpha\zeta) - \log\det M_3([1 - \alpha]\xi + \alpha\zeta)$. Let P be a $k \times k$ matrix so that $PP' = I$ and $PM(\zeta)P'$ is a diagonal matrix with diagonal elements d_i. Let Q be an $r \times r$ matrix so that $QQ' = I$ and $QM_3(\zeta)Q'$ is a diagonal matrix with diagonal elements e_i. (Note, $\operatorname{tr} M(\zeta) = \sum d_i$, and $\operatorname{tr} M_3(\zeta) = \sum e_i$.) Then, again recalling that $M(\xi) = I$,

$$\begin{aligned}q(\alpha) = &\textstyle\sum_1^k \log(1 - \alpha + \alpha d_i) + \log\det PP' - \sum_1^r \log(1 - \alpha + \alpha e_i) \\ &- \log\det QQ'\end{aligned}$$

and

$$q(0) = 0.$$

The first derivative is

$$q'(\alpha) = \textstyle\sum_1^k (-1 + d_i)/(1 - \alpha + \alpha d_i) - \sum_1^r (-1 + e_i)/(1 - \alpha + \alpha e_i)$$

and

$$\begin{aligned}q'(0) &= \operatorname{tr} M_1(\zeta) - s \\ &= \epsilon.\end{aligned}$$

Differentiating again gives

$$\begin{aligned}q''(\alpha) &= -\textstyle\sum_1^k (1 - d_i)^2/(1 - \alpha + \alpha d_i)^2 + \sum_1^r (1 - e_i)^2/(1 - \alpha + \alpha e_i)^2 \\ &\geqq -\textstyle\sum_1^k (1 - d_i)^2/(1 - \alpha + \alpha d_i)^2 \\ &= p(\alpha),\end{aligned}$$

defining $p(\alpha)$. For $0 \leqq \alpha < 1$, $p''(\alpha) \leqq 0$. Therefore for any interval $0 \leqq \alpha \leqq b < 1$, $p(\alpha)$ attains its minimum value at 0 or b. Thus

$$q''(\alpha) \geqq -\max\{\sum (d_i - 1)^2, \sum (d_i - 1)^2/(1 - b + b\,d_i)^2\} \quad \text{for} \quad 0 \leqq \alpha \leqq b < 1. \tag{4.8}$$

Now $\sum (d_i - 1)^2$ is convex in the d_i on the set $B = \{(d_1, \cdots, d_k) \mid \text{all } d_i \geqq 0, \sum d_i \leqq c\}$. This set contains the $(d_1, \cdots, d_k)$ corresponding to ζ, since $\sum d_i = \operatorname{tr} M(\zeta) = \int d(x, \xi)\, d\zeta \leqq c$. Therefore $\sum (d_i - 1)^2$ takes its maximum value if all but one d_i are 0 and the one equals c. Thus

$$\sum (d_i - 1)^2 \leqq k - 1 + (c - 1)^2.$$

To bound the second term on the right in (4.8) let $h(u) = (u - 1)^2/(1 - b + bu)^2$. Then $h''(u) = 2(1 + 2b - 2bu)(1 - b + bu)^{-4}$, which is non-negative if $0 \leqq u \leqq 1 + \frac{1}{2}b^{-1}$. Since $d_i \leqq c$ we conclude that if $b = \frac{1}{2}(c - 1)^{-1}$, (which is less than 1, as required, because $c > \frac{3}{2}$), then the second expression on the right side of (4.8) is convex in the d_i on the set B. Therefore the expression attains its maximum if one d_i equals c and all the rest are 0. Thus

$$\begin{aligned}\sum (d_i - 1)^2/(1 - b + b\,d_i)^2 &\leqq (k - 1)/(1 - b)^2 + (c - 1)^2/(1 + b(c - 1))^2 \\ &= 4(k - 1)(c - 1)^2/(2c - 3)^2 + 4(c - 1)^2/9\end{aligned}$$

where we have replaced b by its value $\frac{1}{2}(c - 1)^{-1}$.

Therefore for $0 \leqq \alpha \leqq \frac{1}{2}(c - 1)^{-1}$,

$$\begin{aligned}q''(\alpha) &\geqq -\max\{k - 1 + (c - 1)^2, 4(k - 1)(c - 1)^2/ \\ &\qquad (2c - 3)^2 + 4(c - 1)^2/9\} \\ &= -K\end{aligned} \tag{4.9}$$

defining K. Therefore for $0 \leqq \alpha \leqq \frac{1}{2}(c - 1)^{-1}$,

$$q(\alpha) \geqq \epsilon\alpha - K\alpha^2/2. \tag{4.10}$$

Now $\epsilon = \bar{d}_{s|k}(\xi) - s \leqq \bar{d}(\xi) - s \leqq c - s$. Thus if $(c - s)/K \leqq \frac{1}{2}(c - 1)^{-1}$ then certainly $\epsilon/K \leqq \frac{1}{2}(c - 1)^{-1}$ and the expression on the right of (4.10) is maximized at $\alpha_0 = \epsilon/K$. Let $\xi' = (1 - \alpha_0)\xi + \alpha_0\zeta$. Then

$$\log[\det M(\xi')/\det M(\xi)] = q(\alpha_0) \geqq \epsilon^2/2K.$$

On the other hand if $(c - s)/K > \frac{1}{2}(c - 1)^{-1}$, then

$$q(\alpha) \geqq \epsilon\alpha - (c - 1)(c - s)\alpha^2,$$

and the right side is maximized at $\alpha_1 = \epsilon/2(c - 1)(c - s)$. In this case let $\xi' = (1 - \alpha_1)\xi + \alpha_1\zeta$. Then $\log[\det M(\xi')/\det M(\xi)] \geqq \epsilon^2/4(c - 1)(c - s)$. This completes the proof.

The bound given by the first statement of the theorem is seen to be sharp if we consider the following example. Let $\mathfrak{X} = \{x_1, \cdots, x_k\} \cup \{y_1, \cdots, y_s\}$, let

$f_i(x_j) = \delta_{ij}$, and let $f_i(y_j) = a\delta_{ij}$, where $|a| > 1$. Then for any design ξ, $M^*(\xi)$ is diagonal with diagonal entries $\xi(x_i) + a^2\xi(y_i)$. Therefore the optimal design η is uniform on $\{y_1, \cdots, y_s\}$. Let ξ_1 be uniform on $\{x_1, \cdots, x_s\}$. Then $\det M^*(\xi_1) = s^{-s}$ and $\det M^*(\eta) = a^{2s}s^{-s}$. Therefore the D-efficiency of ξ_1 is a^{-2}. But $d_{s|k}(x, \xi_1)$ is maximized if x is some y_i. So $\bar{d}_{s|k}(\xi_1) = sa^2$, and $s/\bar{d}_{s|k}(\xi_1)$ is also a^{-2}.

The bound given by the second assertion of the theorem is sharp only in the trivial case when $\bar{d}_{s|k}(\xi) = s$. For if $\bar{d}_{s|k}(\xi) > s$ then $\epsilon > 0$ and hence α_0 and α_1 are nonzero (by definition). If the bound of the theorem is to be sharp, (4.10) must be equality at α_0 (respectively α_1). But this can happen only if $q''(\alpha) = -K$ for all α between 0 and α_0 (respectively α_1), because $-K$ is a lower bound on $q''(\alpha)$. But then $p(\alpha) = -\sum (d_i - 1)^2/(1 - \alpha + \alpha\, d_i)^2$ is constant, $-K$, on the interval, and hence

$$p''(\alpha) = -6 \sum (d_i - 1)^4/(1 - \alpha + \alpha\, d_i)^4 = 0.$$

This is possible only if $d_i = 1$ for every i, in which case $p(\alpha) = 0 \neq -K$. Therefore (4.10) cannot be equality, and the bound in question cannot be sharp when $\bar{d}_{s|k}(\xi) > s$.

We remark that we can use Theorems 4.2 and 4.3 together to get a lower bound on the D-efficiency of the best k-point design for estimating θ. This bound is $1/k$, which is not as good as the bound of Theorem 4.1.

Theorem 4.3 gives some justification to examination of $d_{s|k}(x, \xi)$ for $s < k$, even when $\bar{d}_{s|k}(\xi) > s$. Although $\bar{d}_{s|k}(\xi)$ does not measure the efficiency of ξ with respect to any intuitively meaningful optimality criterion, it does give an indication of the D-efficiency of ξ.

There is an easy extension of the first part of Theorem 4.3 to the case in which $M(\xi)$ is singular. We state this as a corollary.

COROLLARY. *Given any designs ξ and ξ_1, and $0 < \epsilon < 1$, write*

$$\xi_\epsilon = (1 - \epsilon)\xi + \epsilon\xi_1 .$$

Suppose $M(\xi_\epsilon)$ is nonsingular, $0 < \epsilon < 1$, and $\lim_{\epsilon\to 0} \bar{d}_{s|k}(\xi_\epsilon)$ exists. Then

$$[\det M^*(\xi)/\max_{\xi'} \det M^*(\xi')]^{1/s} \geqq s/\lim_{\epsilon\to 0} \bar{d}_{s|k}(\xi_\epsilon).$$

PROOF. For any ϵ between 0 and 1

$$[\det M^*(\xi_\epsilon)/\max_{\xi'} \det M^*(\xi')]^{1/s} \geqq s/\bar{d}_{s|k}(\xi_\epsilon).$$

Taking the limit in ϵ on each side gives the desired result.

We remark that for a given ξ the value of $\lim_\epsilon \bar{d}_{s|k}(\xi_\epsilon)$ will depend in general on what design ξ_1 is used.

The second assertion of Theorem 4.3 cannot be extended in this manner to the case when $M(\xi)$ is singular, as Example 3.1 shows. There we had ξ^* optimal. If ξ_1 was suitably chosen and ξ_ϵ defined as $(1 - \epsilon)\xi^* + \epsilon\xi_1$, we had

$$\lim_\epsilon \bar{d}_{s|k}(x, \xi_\epsilon) = c^2,$$

for arbitrarily large c. On the other hand $\det M^*(\xi_\epsilon)/\det M^*(\xi^*) = 1 - \epsilon$.

Let us now leave the problem of relating D- and G-efficiency. A similar problem is to obtain bounds on the efficiency of a single design in various related models.

THEOREM 4.4. *Suppose ξ_0 is optimal for θ. Then the following are true.*

$$[\det M_1(\xi_0)/\max_\xi \det M_1(\xi)]^{1/s} \geqq s/k. \tag{4.11}$$

$$[\det M^*(\xi_0)/\max_\xi \det M^*(\xi)]^{1/s} \geqq s/k. \tag{4.12}$$

$$d_{s|s}(x, \xi_0) \leqq k. \tag{4.13}$$

$$d_{s|k}(x, \xi_0) \leqq k. \tag{4.14}$$

Statements (4.11) and (4.12) say that ξ_0 has D-efficiency $\geqq s/k$ for estimating $\theta^{(1)}$ out of $\theta^{(1)}$ or $\theta^{(1)}$ out of θ. Line (4.13) says that ξ_0 has G-efficiency $\geqq s/k$ when estimating $f'(x)\theta$. Line (4.14) does not have a natural efficiency interpretation except to the extent given by Theorem 4.3, but it is included here for the sake of completeness.

PROOF. Statements (4.11) and (4.12) can be proved directly, but they follow immediately from (4.13) and (4.14) respectively, by Theorem 4.3. So we need only show (4.13) and (4.14).

We will suppress the ξ_0, writing M_1 for $M_1(\xi_0)$, etc. Let

$$L = \begin{Vmatrix} I & 0 \\ -M_2'M_1^{-1} & I \end{Vmatrix}.$$

Then

$$LML' = \begin{Vmatrix} M_1 & 0 \\ 0 & M_3 - M_2'M_1^{-1}M_2 \end{Vmatrix}.$$

Write $g = Lf$, and note that $g^{(1)} = f^{(1)}$. Then

$$\begin{aligned} k \geqq d_{k|k}(x, \xi_0) &= g'(x)(LML')^{-1}g(x) \\ &= g^{(1)\prime}(x)M_1^{-1}g^{(1)}(x) + g^{(2)\prime}(x)[M_3 - M_2'M_1^{-1}M_2]^{-1}g^{(2)}(x) \\ &\geqq g^{(1)\prime}(x)M_1^{-1}g^{(1)}(x) \\ &= f^{(1)\prime}(x)M_1^{-1}f^{(1)}(x) \\ &= d_{s|s}(x, \xi_0), \end{aligned}$$

proving (4.13).

To prove (4.14), write

$$\begin{aligned} d_{s|k}(x, \xi_0) &= f'(x)Mf(x) - f^{(2)\prime}(x)M_3^{-1}f^{(2)}(x) \\ &\leqq f'(x)Mf(x) \\ &= d_{k|k}(x, \xi_0) \leqq k, \end{aligned}$$

as was to be shown.

The bounds are sharp, as is shown by the following example in which each bound is attained. Note that in this example the inequalities in the proof are equality at $x_1, \cdots, x_s$, and the arithmetic-geometric mean inequality, used in the proof of Theorem 4.3, is equality.

Let $\mathfrak{X} = \{x_1, \cdots, x_k\}, f_i(x_j) = \delta_{ij}$. Then ξ_0 is uniform on $\mathfrak{X}$. For a competing design in (4.11) and (4.12) let ξ be uniform on $\{x_1, \cdots, x_s\}$. It is immediate that all four bounds in the theorem are attained.

We now apply this result to the product situation considered by Hoel (1965). Suppose the regression functions $f_{ij}(y, z)$ on $\mathcal{Y} \times \mathcal{Z}$ can be factored into $f_{ij}(y, z) = g_i(y)h_j(z)$. The optimal design for the coefficients of the g_i alone is some measure η on $\mathcal{Y}$, and for the coefficients of the h_j a measure ζ on $\mathcal{Z}$. Hoel then shows that an optimal design for the coefficients of the f_{ij} is simply $\eta \times \zeta$. This gives a corollary to Theorem 4.4.

COROLLARY. *Let η and ζ be optimal designs for the coefficients of some n-vectors of regression functions g and h on spaces $\mathcal{Y}$ and $\mathcal{Z}$ respectively. If $f(y, z)$ consists of some s of the products $g_i(y)h_j(z)$, then for estimating the coefficients of f the product design $\eta \times \zeta$ has D-efficiency $\geqq s/n^2$ and G-efficiency $\geqq s/n^2$.*

For example if $\mathcal{Y} = \mathcal{Z} = [-1, 1]$ and $g(y)$ and $h(z)$ each consist of the n monomials of degree $\leqq n - 1$, then the optimal design in each one dimensional problem is well known. See Guest (1958) and Hoel (1958). If $f(y, z)$ consists of those monomials in y and z of total degree $\leqq n - 1$, then the product design ξ_0 has D-efficiency $\geqq (n + 1)/2n$ and G-efficiency $\geqq (n + 1)/2n$.

In this situation an optimal design is known only for small n. (See Kiefer (1961, Section 4) and Farrell, Kiefer and Walbran (1965, Section 3).) Even though ξ_0 is not optimal it is easy to obtain and use, so a bound such as that given, showing that ξ_0 is fairly efficient, is nice to have.

In fact in this example of polynomial regression, ξ_0 seems to be considerably more efficient than the bound indicates. If f is quadratic regression on the square (i.e., $n = 3$), Kiefer (1961, pages 314–317) has computed an optimal design ξ^* for the six regression coefficients. We compare this with the product design ξ_0. (M here is a 6×6 matrix.)

$$[\det M(\xi_0)]^{1/6} = .462$$

$$[\det M(\xi^*)]^{1/6} = .475$$

$$D\text{-efficiency of } \xi_0 = .97$$

$$\text{Theoretical lower bound} = \tfrac{6}{9} = .67$$

$$\max_x d_{6|6}(x, \xi_0) = 7.25$$

$$G\text{-efficiency of } \xi_0 = .83$$

$$\text{Theoretical lower bound} = \tfrac{6}{9} = .67.$$

Although in this example the efficiencies are greater than the lower bounds of Theorem 4.4, there are product design situations in which the bounds are attained. As an example suppose $\mathcal{Y} = \{y_1, \cdots, y_m\}, \mathcal{Z} = \{z_1, \cdots, z_n\}$, $g(y)$ is an

m-vector with $g_i(y_j) = \delta_{ij}$, and $h(z)$ is an n-vector with $h_i(z_j) = \delta_{ij}$. Let the regression functions $f_l(y, z)$ comprise some s-element subset of the mn functions $g_i(y)h_j(z)$. Then $f(y, z)$ is just that given in the example immediately following Theorem 4.4, in which all the lower bounds on efficiency were attained.

Acknowledgment. I wish to express my deepest appreciation to Professor Jack Kiefer for his interest and helpful guidance while this paper was being written.

REFERENCES

[1] Atwood, C. L. (1968). Optimal and efficient designs of experiments. Dept. of Statistics Technical Report No. 106, Univ. of Minnesota.

[2] Box, G. E. P. and Lucas, H. L. (1959). Design of experiments in non-linear situations. *Biometrika* **46** 77–90.

[3] Box, M. J. (1968). The occurrence of replications in optimal designs of experiments to estimate parameters in non-linear models. *J. Roy. Statist. Soc. Ser. B* **30** 290–302.

[4] Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24** 586–602.

[5] Farrell, R. H., Kiefer, J. and Walbran, A. (1965). Optimum multivariate designs. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 113–138. Univ. of California Press.

[6] Guest, P. G. (1958). The spacing of observations in polynomial regression. *Ann. Math. Statist.* **29** 294–299.

[7] Hoel, P. G. (1958). Efficiency problems in polynomial estimation. *Ann. Math. Statist.* **29** 1134–1145.

[8] Hoel, P. G. (1965). Minimax designs in two dimensional regression. *Ann. Math. Statist.* **36** 1097–1106.

[9] Householder, A. S. (1964). *The Theory of Matrices in Numerical Analysis*. Blaisdell, Waltham.

[10] Karlin, S. and Studden, W. J. (1966a). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815.

[11] Karlin, S. and Studden, W. J. (1966b). *Tchebycheff Systems: with Applications in Analysis and Statistics*. Interscience, New York.

[12] Keilson, J. (1966). A theorem on optimum allocation for a class of symmetric multilinear return functions. *J. Math. Anal. Appl.* **15** 269–272.

[13] Kiefer, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **21** 272–304.

[14] Kiefer, J. (1960). Optimum experimental designs V, with applications to systematic and rotatable designs. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 381–405. Univ. of California Press.

[15] Kiefer, J. (1961). Optimum designs in regression problems, II. *Ann. Math. Statist.* **32** 298–325.

[16] Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

[17] Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **14** 363–366.

[18] Meeter, D. (1967). Mimeographed class notes. Florida State University.

[19] Scheffé, H. (1958). Experiments with mixtures. *J. Roy. Statist. Soc. Ser. B* **20** 344–360.

[20] Stone, M. (1959). Application of a measure of information to the design and comparison of regression experiments. *Ann. Math. Statist.* **30** 55–70.

[21] Uranisi, H. (1964). Optimum design for the special cubic regression on the q-simplex. *Mathematical Reports* **1** 7–12. General Ed. Dept., Kyushu University.