The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1435–1447

OPTIMAL DESIGNS ON TCHEBYCHEFF POINTS[1]

BY W. J. STUDDEN

Purdue University

0. Summary. Kiefer and Wolfowitz (1959) proved that the optimal design for estimating the highest coefficient in polynomial regression is supported by certain Tchebycheff points. Hoel and Levine (1964) showed that the optimal designs for extrapolation in polynomial regression were all supported by the Tchebycheff points. These results were extended by Kiefer and Wolfowitz (1965) to cover nonpolynomial regression problems involving Tchebycheff systems and the large class of designs supported by the Tchebycheff points was characterized. In the present paper it is shown that the optimal design for estimating any specific parameter is supported by one of two sets of points for Tchebycheff systems with certain symmetry properties. Different proofs of the Kiefer-Wolfowitz results are also presented. The author wishes to thank Professor Kiefer for providing one of the counterexamples in Section 6.

1. Introduction. Let $f = (f_0, f_1, \cdots, f_n)$ denote $n + 1$ linearly independent continuous functions on a compact set $\mathfrak{X}$. For each $x \varepsilon \mathfrak{X}$ an experiment can be performed. The outcome is a random variable $y(x)$ with mean value $\sum_{i=0}^n \theta_i f_i(x)$ and a common variance σ^2. The functions $f_0, f_1, \cdots, f_n$, called the regression functions, are assumed known while $\theta_0, \theta_1, \cdots, \theta_n$ and σ^2 are unknown. An experimental design is a probability measure ξ on $\mathfrak{X}$. The problem concerned with here is that of estimating a linear form $(c, \theta) = \sum_{i=0}^n c_i\theta_i$. It will always be assumed that $\sum_{i=0}^n c_i^2 > 0$. For a given design ξ let $m_{ij} = m_{ij}(\xi) = \int f_i f_j \, d\xi$ and $M(\xi) = \|m_{ij}(\xi)\|_{i,j=0}^n$. A linear form (c, θ) is called estimable with respect to ξ if c is contained in the range of the matrix $M(\xi)$. If c is estimable with respect to ξ let

$$V(c, \xi) = \sup\, (c, d)^2/(d, M(\xi)\, d)$$

where the sup is taken over the set of vectors d such that the denominator is nonzero. If c is not estimable with respect to ξ we define $V(c, \xi) = \infty$. Suppose ξ concentrates mass ξ_i at the points x_i, $i = 1, 2, \cdots, r$, and $\xi_i N = n_i$ are integers. If N uncorrelated observations are made, taking $N\xi_i$ observations at x_i, then the variance of the best linear unbiased estimate of (c, θ) is given by $\sigma^2 N^{-1} V(c, \xi)$. An arbitrary measure or design ξ is called *c-optimal* if ξ minimizes $V(c, \xi)$. For a more complete discussion of the above model see Kiefer (1959) or Karlin and Studden (1966b).

Received 21 September 1967.

[1]This research was supported in part by the Office of Naval Research Contract NONR 1100(26) and the Aerospace Research Laboratories Contract AF 33(615)67C1244 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

We assume throughout that $\mathfrak{X} = [-1, 1]$. Hoel and Levine showed that if $f_i(x) = x^i, i = 0, 1, \cdots, n$, and $c = f(x_0)$ with $|x_0| > 1$ then the c-optimal design is supported on the Tchebycheff points $s_\nu = -\cos(\nu\pi/n)$, $\nu = 0, 1, \cdots, n$. These are the points where $|T_n(x)| = 1$, $T_n(x)$ being the nth Tchebycheff polynomial of the 1st kind.

Kiefer and Wolfowitz consider more general systems of regression of functions and a related set of Tchebycheff points. Let T^* denote the set of all c such that a c-optimal design is supported on the entire set of Tchebycheff points. The set T^* is divided into 2 different sets R^* and $S^* = T^* - R^*$ which they explicitly characterize. Moreover they show that the set R^* includes the set A^* of all vectors c for which (c, θ) is not estimable for any design on fewer that $n + 1$ points. The set A^* may be characterized by stipulating that $c \varepsilon A^*$ if and only if the determinants

$$\begin{vmatrix} f_0(x_1) & \cdots & f_0(x_n) & c_0 \\ f_1(x_1) & \cdots & f_1(x_n) & c_1 \\ \vdots & & \vdots & \vdots \\ f_n(x_1) & \cdots & f_n(x_n) & c_n \end{vmatrix} \tag{1.1}$$

do not vanish whenever the x_i are all distinct.

In Sections 2 and 3 we offer different proofs of the Kiefer-Wolfowitz results and more emphasis is placed on the fact that their sets S^* and R^* can be characterized in a manner similar to the determinant characterization of A^*. More explicitly the set R^* is precisely the set of c for which the determinants (1.1) are of *one strict* sign for $x_1, x_2, \cdots, x_n$ an ordered subset of the Tchebycheff points $s_0 < s_1 < \cdots < s_n$ and S^* consists of those vectors c for which the determinants (1.1) *alternate* in sign as we progressively omit successive s_i. The proofs offered here rely heavily on the elegant result of Elfving which is stated in Theorem 2.1. A close inspection of the analysis used by Kiefer and Wolfowitz and the analysis used here shows certain similarities, however much of the preliminary discussion and game theory has been eliminated using the Elfving result.

In Section 4 it is shown that the vectors $c_p = (0, \cdots, 0, 1, 0, \cdots, 0)$ (a one only in the $p + 1$st component) are in R^* if $n - p$ is even while if $n - p$ is odd the c_p-optimal design is supported on the Tchebycheff points of one lower order. The case $p = n$ was originally proven in Kiefer and Wolfowitz (1959). The above shows that in the case of ordinary polynomial regression the optimum design for estimating θ_p is supported by the set $s_\nu = -\cos(\nu\pi/n)$, $\nu = 0, 1, \cdots, n$, when $n - p$ is even and by the set $t_\nu = -\cos(\nu\pi/(n-1))$, $\nu = 0, 1, \cdots, n-1$, when $n - p$ is odd.

Section 5 contains some additional remarks which show that certain linear combinations of the vector c_p are also supported by the Tchebycheff points while Section 6 contains a simple counterexample which shows the minimax design is not necessarily supported by the minimal number of $n + 1$ points when $f_0, f_1, \cdots, f_n$ form a Tchebycheff system.

2. Designs supported by Tchebycheff points. The following result due to Elfving (1952) characterizes the c-optimal designs ξ and will be frequently employed throughout the paper.

THEOREM 2.1. *Let* $\mathcal{R}_+ = \{f(x) = (f_0(x), \cdots, f_n(x)) \mid x \varepsilon \mathfrak{X}\}$, $\mathcal{R}_- = \{-f(x) \mid x \varepsilon \mathfrak{X}\}$ *and* $\mathcal{R}$ = *the convex hull of* $\mathcal{R}_+ \cup \mathcal{R}_-$. *A design* ξ_0 *is c-optimum if and only if there exists a measurable function* $\varphi(x)$ *satisfying* $|\varphi(x)| \equiv 1$ *such that* (i) $\int \varphi(x) f(x) \xi(dx) = \beta c$ *for some* β *and* (ii) βc *is a boundary point of* $\mathcal{R}$. *Moreover* βc *lies on the boundary of* $\mathcal{R}$ *if and only if* $\beta^2 = v_0^{-1}$ *where* $v_0 = \min_\xi V(c, \xi)$.

Every vector $c \varepsilon \mathcal{R}$ can be put in the form

$$c = \textstyle\sum_{\nu=1}^{k} \epsilon_\nu p_\nu f(x_\nu) \tag{2.1}$$

where $\epsilon_\nu = \pm 1$, $p_\nu > 0$ and $\sum_1^k p_\nu = 1$. The integer k may always be taken to be at most $n + 2$ and at most $n + 1$ if c is a boundary point of $\mathcal{R}$. The following simple lemma will be needed.

LEMMA 2.1. *A vector c of the form* (2.1) *lies on the boundary of* $\mathcal{R}$ *if and only if there exists a nontrivial "polynomial"* $u(x) = \sum_\nu a_\nu f_\nu(x)$ *such that* $|u(x)| \leqq 1$ *for* $x \varepsilon [-1, 1]$, $\epsilon_\nu u(x_\nu) = 1$, $\nu = 1, 2, \cdots, k$, *and* $\sum_\nu a_\nu c_\nu = 1$.

PROOF. If the required polynomial exists and $a = (a_0, a_1, \cdots, a_n)$ then $(a, c) = \sum_\nu \epsilon_\nu p_\nu (a, f(x_\nu)) = \sum_\nu p_\nu \epsilon_\nu u(x_\nu) = 1$ and $(a, y) \leqq 1$ for all $y \varepsilon \mathcal{R}$. The vector a defines a supporting plane to $\mathcal{R}$ at c so that c is a boundary point of $\mathcal{R}$.

If c is a boundary point of $\mathcal{R}$ then a supporting plane exists, i.e. there exists a vector $a \neq 0$ such that $(a, c) = 1$ and $(a, y) \leqq 1$ for all $y \varepsilon \mathcal{R}$. (Note that the origin is in the interior of $\mathcal{R}$.) Therefore $1 = (a, c) = \sum_{\nu=1}^k p_\nu \epsilon_\nu (a, f(x_\nu))$ and $|(a, f(x))| = |u(x)| \leqq 1$ for all $x \varepsilon [-1, 1]$. In this case $\epsilon_\nu(a, f(x_\nu)) = 1$ for $\nu = 1, \cdots, k$ since we have assumed $p_\nu > 0$.

REMARK 2.1. For an arbitrary vector $c \neq (0, \cdots, 0)$, βc lies on the boundary of $\mathcal{R}$ for some $\beta > 0$ and hence $\beta c = \sum_{\nu=1}^{n+1} \epsilon_\nu p_\nu f(x_\nu)$ for some $\{\epsilon_\nu p_\nu\}$ and $\{x_\nu\}$. If $(a, f) = \sum_{i=0}^n a_i f_i$ denotes the polynomial of Lemma 2.1 then the minimal value of $V(c, \xi)$ is $\beta^{-2} = (\sum_0^n a_i c_i)^2 = (a, c)^2$ since $(\beta c, a) = 1$. Moreover

$$\begin{aligned}\inf_\xi V(c, \xi) &= \inf_\xi \sup_b (c, b)^2 [\textstyle\int (b, f(x))^2 \xi(dx)]^{-1} \\ &\geqq \sup_b \inf_\xi (c, b)^2 [\textstyle\int (b, f(x))^2 \xi(dx)]^{-1} \\ &\geqq (c, a)^2.\end{aligned}$$

Since the first and last terms are equal

$$\inf_\xi \sup_b (c, b)^2 [\textstyle\int (b, f(x))^2 \xi(dx)]^{-1} = \sup_b \inf_\xi (c, b)^2 [\textstyle\int (b, f(x))^2 \xi(dx)]^{-1}.$$

We shall assume throughout the paper that the set of functions $f_0, f_1, \cdots, f_n$ form a Tchebycheff system or a T-system on $[-1, 1]$ and that $U(x) = \sum_i a_i f_i(x) \equiv 1$ for some set $\{a_i\}_0^n$. A T-system $f_0, f_1, \cdots, f_n$ is such that every linear combination $\sum_i a_i f_i(x) [\sum_0^n a_i^2 > 0]$ has at most n distinct zeros on $[-1, 1]$ or equivalently the determinants

$$(2.2)\qquad F\begin{pmatrix}0, 1, \cdots, n\\ x_0, x_1, \cdots, x_n\end{pmatrix} = \begin{vmatrix} f_0(y_0) & f_0(x_1) & \cdots & f_0(x_n)\\ f_1(y_0) & f_1(x_1) & \cdots & f_1(x_n)\\ \vdots & \vdots & & \vdots\\ f_n(x_0) & f_n(x_1) & \cdots & f_n(x_n)\end{vmatrix}$$

are of one strict sign provided $-1 \leqq x_0 < x_1 < \cdots < x_n \leqq 1$. For definiteness we shall assume that the determinants are positive. We shall call any linear combination $\sum_{i=0}^n a_i f_i(x)$ a polynomial. The classical example of a T-system consists of the ordinary powers $f_i(x) = x^i$, $i = 0, 1, \cdots, n$. In this case the determinant (2.2) reduces to the classical Vandermonde determinant. A simple constructive method of generating more general T-systems will be given in Section 4.

The following important property of T-systems will be used (see Karlin and Studden (1966a), Theorem II. 10.1). If $\{f_i\}_0^n$ is a T-system on $[-1, 1]$ then there exists a unique polynomial $W(x) = \sum_{i=0}^n a_i^* f_i(x)$ satisfying the properties

(i) $|W(x)| \leqq 1$,

(ii) there exists $n + 1$ points $-1 \leqq s_0 < s_1 < \cdots < s_n \leqq 1$ such that $W(s_i) = (-1)^{n-i}$, $i = 0, 1, \cdots, n$. Moreover when $U(x) \equiv 1$ is a polynomial equality occurs in (i) *only* for $x = s_0, s_1, \cdots, s_n$ and $s_0 = -1$ and $s_n = +1$.

For any vector $c(\neq (0, \cdots, 0))$ considerable use will be made of the determinants

$$(2.3)\qquad D_\nu(c) = \begin{vmatrix} f_0(s_0) & \cdots & f_0(s_{\nu-1}) & f_0(s_{\nu+1}) & \cdots & f_0(s_n) & c_0\\ f_1(s_0) & \cdots & f_1(s_{\nu-1}) & f_1(s_{\nu+1}) & \cdots & f_1(s_n) & c_1\\ \vdots & & \vdots & \vdots & & \vdots & \vdots\\ f_n(s_0) & \cdots & f_n(s_{\nu-1}) & f_n(s_{\nu+1}) & \cdots & f_n(s_n) & c_n\end{vmatrix}.$$

The sign of $D_\nu(c)$ will be denoted by $d_\nu(c)$; if $D_\nu(c) = 0$ the sign may be defined as -1 or $+1$. We further let $L_\nu(x) = \sum_i a_{\nu i} f_i(x)$ denote the Lagrange interpolation polynomial defined by requiring that $L_\nu(s_j) = \delta_{\nu j}$, $\nu, j = 0, 1, \cdots, n$. In terms of the determinants (2.2) $L_\nu(x)$ has the more explicit form

$$L_\nu(x) = F\begin{pmatrix}0, \cdots, \nu-1, \nu, \nu+1, \cdots, n\\ s_0, \cdots, s_{\nu-1}, x, s_{\nu+1}, \cdots, s_n\end{pmatrix} \Big/ F\begin{pmatrix}0, \cdots, n\\ s_0, \cdots, s_n\end{pmatrix}.$$

For any vector c and any polynomial $u(x) = \sum_i a_i f_i(x)$ we shall use the notation $u^*(c) = \sum_i a_i c_i$. In this case we have

$$(-1)^{n-\nu} D_\nu(c) = L_\nu^*(c) F\begin{pmatrix}0, \cdots, n\\ s_0, \cdots, s_n\end{pmatrix}.$$

Now for any polynomial $u(x)$ we have $u(x) = \sum_{\nu=0}^n u(s_\nu) L_\nu(x)$. Since the coefficients of f_i on either side are equal we find that $u^*(c) = \sum_{\nu=0}^n u(s_\nu) L_\nu^*(c)$. Letting $u(x)$ be successively the polynomials $f_0, f_1, \cdots, f_n$ we find that

$$\begin{aligned} c &= \textstyle\sum_{\nu=0}^n f(s_\nu) L_\nu^*(c)\\ &= \textstyle\sum_{\nu=0}^n (-1)^{n-\nu} d_\nu(c) |L_\nu^*(c)| f(s_\nu).\end{aligned}$$

Therefore

$$\beta c = \sum_{\nu=0}^{n} (-1)^{n-\nu} d_\nu(c) p_\nu f(s_\nu) \tag{2.4}$$

where

$$p_\nu = |L_\nu^*(c)|/\sum_{j=0}^{n} |L_j^*(c)| = |D_\nu(c)|/\sum_{j=0}^{n} |D_j(c)|, \qquad \nu = 0, 1, \cdots, n,$$

and

$$\beta = (\sum_{j=0}^{n} |L_j^*(c)|)^{-1} = F\begin{pmatrix} 0, & \cdots, & n \\ s_0, & \cdots, & s_n \end{pmatrix} (\sum_{j=0}^{n} |D_j(c)|)^{-1}.$$

Let R denote the class of vectors c such that $\epsilon D_\nu(c) \geqq 0$ for $\nu = 0, 1, \cdots, n$ where ϵ is fixed to be $+1$ or -1 for a given vector c (i.e. the $D_\nu(c)$, $\nu = 0, 1, \cdots, n$, all have the same sign in a weak sense) and let S denote those c for which $\epsilon(-1)^\nu D_\nu(c) \leqq 0$, $\nu = 0, 1, \cdots, n$. The following theorem is a slight generalization of Theorem 3.1 of Karlin and Studden (1966) which considers only those vectors c contained in the set A^* which is described in the next section. The theorem is also very closely related to the results of Kiefer and Wolfowitz (1965) which are discussed in the next section.

THEOREM 2.2. *Suppose that $\{f_i\}_0^n$ is a T-system such that $U(x) \equiv 1$ is a polynomial.*

(a) *For any design ξ*

$$\begin{aligned} d(c, \xi) &\geqq [W^*(c)]^2, \qquad c \in R, \\ &\geqq [U^*(c)]^2, \qquad c \in S, \end{aligned} \tag{2.5}$$

where $W(x)$ is the oscillatory polynomial defined above.

(b) *Equality occurs in* (2.5) *for $\xi = \xi_0$ concentrating mass*

$$p_\nu = |L_\nu^*(c)|/\sum_{\nu=0}^{n} |L_\nu^*(c)| = |D_\nu(c)|/\sum_{\nu=0}^{n} |D_\nu(c)|$$

at the points s_ν, $\nu = 0, 1, \cdots, n$.

(c) *The design ξ_0 is the only design supported on $s_0 < \cdots < s_n$ attaining equality in* (2.5). *If $c \in R$ then ξ_0 is the only design attaining equality in* (2.5).

PROOF. (a) Using the polynomials W and U in Lemma 2.1 we find that for $c \in R \cup S$ the quantity βc in (2.4) is a boundary point of $\mathcal{R}$. Moreover for $c \in R$, $\beta W^*(c) = 1$ so that the minimum value of $V(c, \xi)$ is $\beta^{-2} = [W^*(c)]^2$. Similarly $[U^*(c)]^2$ is the minimal value of $V(c, \xi)$ for $c \in S$.

(b) The design ξ_0 is c-optimal in each case by Lemma 2.1 and the properties of U and W.

(c) For $c \in S$ and any c-optimal design supported on $s_0, s_1, \cdots, s_n$, equation (2.4) holds for some set $(-1)^{n-\nu} d_\nu(c) p_\nu$, $\nu = 0, 1, \cdots, n$, where $\beta^{-2} = (U^*(c))^2$. However this set of linear equations has a unique solution since $F\binom{0, \cdots, n}{s_0, \cdots, s_n} \neq 0$. For $c \in R$ we have

$$V(c, \xi) = \sup_d [(c, d)^2/(d, M(\xi)\, d)] \geqq (W^*(c))^2/\int (W(x))^2 d\xi \geqq (W^*(c))^2.$$

The last inequality is strict unless ξ is supported on $s_0, s_1, \cdots, s_n$ since $|W(x)| < 1$

for $x \neq s_0, s_1, \cdots, s_n$. The last sentence in (c) now follows since the c-optimum design supported by $s_0, s_1, \cdots, s_n$ is again unique.

REMARK 2.2. As observed by Kiefer and Wolfowitz the above theorem is a consequence of the fact that the convex hulls A and B of the two sets $\{f(s_\nu)\}_0^n$ and $\{(-1)^{n-\nu}f(s_\nu)\}_0^n$ both lie in boundary faces of the set $\mathcal{R}$ defined by Theorem 2.1. It is easily seen that B is the *entire* face since $|W(x)| = 1$ only for $x = s_0, s_1, \cdots, s_n$, i.e. the intersection of $\mathcal{R}$ and the hyperplane defined by $W^*(c) = 1$ is precisely B. The union of B and its symmetric image is the set R.

The set A is properly contained in the face of $\mathcal{R}$ determined by $U^*(c) = 1$. The complete face in this case is the convex hull of the whole curve $\{f(x) \mid x \in [-1, 1]\}$ because of the assumption that $U(x) \equiv 1$ is a polynomial; thus every vector $c = \sum_\nu p_\nu f(x_\nu)$ is a boundary point of $\mathcal{R}$. The usefulness of S will be apparent in the next section.

3. The Kiefer-Wolfowitz theorems. Kiefer and Wolfowitz (1965) define five sets of vectors T^*, R^*, S^*, A^*, H^*. The first set T^* consists of those c such that a c-optimal design exists on the *full* set of Tchebycheff points $s_0, s_1, \cdots, s_n$. The next two sets R^* and S^* consist of those c which for some $\beta \neq 0$ are of the form

$$\beta c = \sum_{\nu=0}^n \epsilon_\nu p_\nu f(s_\nu), \quad p_\nu > 0, \quad \sum_0^n p_\nu = 1$$

where the ϵ_ν alternate in sign for $c \in R^*$ while the ϵ_ν are the same sign for the vectors in S^*. The set A^* consists of those c for which

$$\begin{vmatrix} f_0(x_1) & \cdots & f_0(x_n) & c_0 \\ f_1(x_1) & \cdots & f_1(x_n) & c_1 \\ \vdots & & \vdots & \vdots \\ f_n(x_1) & \cdots & f_n(x_n) & c_n \end{vmatrix} \neq 0, \tag{3.1}$$

provided the x_i are distinct and lie in $[-1, 1]$. The set A^* is shown to consist of those c such that (c, θ) is only estimable for designs supported by at least $n + 1$ points. Finally to define H^* it is assumed that the regression functions $f_0, f_1, \cdots, f_n$ are defined on some interval containing $[-1, 1]$ and $c \in H^*$ if and only if $c = \beta f(x_0)$ for some $|x_0| > 1$ and $\beta \neq 0$.

With suitable further assumptions (see below) on $f_0, f_1, \cdots, f_n$ it is shown that:

(i) R^* contains a neighborhood of $c = (0, \cdots, 0, 1)$.

(ii) $H^* \subset A^* \subset R^*$.

(iii) The vectors $c \in R^*$ have unique optimal designs while the vectors $c \in S^*$ have unique designs among those supported by $s_0, \cdots, s_n$.

(iv) R^* and S^* are disjoint and $R^* \cup S^* = T^*$.

In Section 2 we have formulated our sets R and S entirely in terms of the determinants (2.3) or (3.1) using only the points $s_0, s_1, \cdots, s_n$. If we define R_0 and S_0 as the subsets of R and S respectively such that the determinants $D_\nu(c) \neq 0, \nu = 0, 1, \cdots, n$, then it can readily be seen that $R_0 = R^*$ and $S_0 = S^*$. Moreover the sets R_0 and S_0 are simply the interiors of the sets R and S respectively. The majority of the results (i)–(iv) above follow from Theorem 2.2.

(i) To prove (i) it suffices to assume that the determinants

$$(3.2) \qquad \begin{vmatrix} f_0(s_0) & \cdots & f_0(s_{\nu-1}) & f_0(s_{\nu+1}) & \cdots & f_0(s_n) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ f_{n-1}(s_0) & \cdots & f_{n-1}(s_{\nu-1}) & f_{n-1}(s_{\nu+1}) & \cdots & f_{n-1}(s_n) \end{vmatrix}$$

for $\nu = 0, 1, \cdots, n$, are all nonzero and of the same sign. This will be true whenever the functions $f_0, f_1, \cdots, f_{n-1}$ form a T-system on $[-1, 1]$.

(ii) That $A^* \subset R^*$ is clear while $H^* \subset A^*$ provided $f_0, f_1, \cdots, f_n$ is a T-system on $(-\infty, \infty)$. In fact $c = \beta f(x_0) \in A^*$ for fixed $x_0 \notin [-1, 1]$ provided $f_0, f_1, \cdots, f_n$ is a T-system on $[-1, 1] \cup \{x_0\}$.

(iii) This follows directly from Theorem 2.2.

(iv) That R^* and S^* are disjoint and both are contained in T^* is immediate once we note their identification with R_0 and S_0. For completeness we will now prove that $R^* \cup S^* = T^*$ (or $R^* \cup S^* \supset T^*$) with the aid of Assumption 2 in Kiefer and Wolfowitz (1965). Assumption 2 states that every polynomial $\sum_0^n a_i f_i(x)$ either has less than or equal to $n - 1$ changes of direction on $(-1, 1)$ or else is constant on $(-1, 1)$. (A function $u(x)$ changes direction at $y \in (-1, 1)$ if $u(x)$ has a local maximum or minimum at y. In particular if $u(x)$ is constant on an open subinterval of $(-1, 1)$ it is said to have infinitely many changes of direction). With this assumption we may prove that $R^* \cup S^* \supset T^*$ (and hence that $R^* \cup S^* = T^*$) as follows. If $c \in T^*$ then

$$\beta c = \sum_{\nu=0}^n \epsilon_\nu p_\nu f(s_\nu)$$

for some $\beta \neq 0$, $e_\nu = \pm 1$, $p_\nu > 0$ and $\sum_0^n p_\nu = 1$. Moreover by Lemma 2.1 there exists a polynomial $u(x) = \sum_0^n a_i f_i(x)$ such that $|u(x)| \leqq 1$ for $x \in [-1, 1]$ and $\epsilon_\nu u(s_\nu) = 1$, $\nu = 0, 1, \cdots, n$. Now if $c \notin R^* \cup S^*$ then the ϵ_ν are not constant (hence $u(x) \not\equiv$ constant) and do not alternate in sign so that there exists a j such that $\epsilon_j \epsilon_{j+1} > 0$. Then $u(x)$ has a change of direction at $s_1, s_2, \cdots, s_{n-1}$ and at least one in the open interval (s_j, s_{j+1}). We therefore have a contradiction since $u(x)$ has n changes of direction and is nonconstant.

4. Optimal designs for the individual regression coefficients. Let $c_p = (0, \cdots, 0, 1, 0, \cdots, 0)$ denote the vector with a one in the $p + 1$st component and zeros elsewhere. In this case $(c_p, \theta) = \theta_p$. For $p = n$ it follows from Theorem 2.2 that if $f_0, \cdots, f_{n-1}$ is a T-system then the unique c_n-optimal design is supported by the full set of Tchebycheff points $s_0 < s_1 < \cdots < s_n$. This result was first proven in Kiefer and Wolfowitz (1959). The purpose of this section is to show that under suitable assumptions the unique c_p-optimal design $(p > 0)$ is supported by this same full set of Tchebycheff points if $n - p$ is even and by the full set of Tchebycheff points of one lower order if $n - p$ is odd. For $p = 0$, $n \geqq 2$ the unique c_p optimal design is supported entirely on $x = 0$. This result is motivated by certain extremal properties of the ordinary Tchebycheff polynomial $T_n(x) = \cos n\theta$, $x = \cos \theta$. The assumptions to be made on the regression functions will be such that they resemble to a very large extent the ordinary powers

$1, x, \cdots, x^n$ on $[-1, 1]$. An example concerning these assumptions is given in Section 6.

We shall assume the following:

(i) $\{f_i\}_0^k$ for $k = n - 2, n - 1, n$, are T-systems on $[-1, 1]$;

(ii) $f_0(x) \equiv 1$;

(iii) $f_i(x) = (-1)^i f_i(-x), i = 0, 1, \cdots, n$;

(iv) for every subset $i_1, i_2, \cdots, i_k$ of $0, 1, \cdots, n$ the system $f_{i_1}(x)$, $f_{i_2}(x), \cdots, f_{i_k}(x)$ is a T-system on the half open interval $(0, 1]$, i.e. every linear combination of $f_{i_1}, \cdots, f_{i_k}$ has at most $k - 1$ distinct zeros in $(0, 1]$;

(v) every polynomial $\sum_0^n a_i f_i$ either has fewer than n changes of direction on $(-1, 1)$ or else is constant on $(-1, 1)$.

Since $\{f_i\}_0^n$ and $\{f_i\}_0^{n-1}$ are both T-systems, the polynomials $W_n(x) = W(x)$ and $W_{n-1}(x)$ exist. (W_{n-1} is defined in terms of $\{f_i\}_0^{n-1}$). Denote the Tchebycheff points for W_n by $s_0 = 1 < s_1 < \cdots < s_n = 1$ as before and the Tchebycheff points for $W_{n-1}(x)$ by $t_0 = -1 < t_1 < \cdots < t_{n-1} = 1$. From the uniqueness properties of W_n and W_{n-1} both of the sets $\{s_i\}_0^n$ and $\{t_i\}_0^{n-1}$ are symmetric about zero as in the case of the ordinary powers. In fact for n even, say $n = 2m$, $W_n = \sum_{k=0}^m a_{2k} f_{2k}$; while for $n = 2m + 1$, $W_n = \sum_{k=0}^m a_{2k+1} f_{2k+1}$.

The powers $1, x, \cdots, x^n$ satisfy all of the conditions (i)–(v). A simple constructive method of obtaining further systems satisfying these conditions is as follows. Let $w_1, w_2, \cdots, w_n$ be any n strictly positive continuous functions on $[-1, 1]$ such that $w_j(x) = w_j(-x)$ for all j and define

$$\begin{aligned}
f_0(x) &= 1, \\
f_1(x) &= \int_0^x w_1(\xi_1)\, d\xi_1, \\
f_2(x) &= \int_0^x w_1(\xi_1) \int_0^{\xi_1} w_2(\xi_2)\, d\xi_2\, d\xi_1, \\
\vdots \quad & \quad \vdots \\
f_n(x) &= \int_0^x w_1(\xi_1) \int_0^{\xi_1} w_2(\xi_2) \cdots \int_0^{\xi_{n-1}} w_n(\xi_n)\, d\xi_n \cdots d\xi_1 .
\end{aligned}$$

It can be shown (see Karlin and Studden (1966a), Chapter 11) that the above system satisfies conditions (i)–(v). The powers arise for the special case $w_i(\xi) \equiv i, i = 1, 2, \cdots, n$.

Now for $p = 0$ and $n = 1$ it is clear that ξ is c_p-optimal if and only if ξ is symmetric about zero. For $p = 0$ and $n \geqq 2$ it is easily seen that the unique c_p-optimum design concentrates mass one at $x = 0$. We shall therefore assume that $p \neq 0$ and $n \geqq 2$.

THEOREM 4.1. *Let* $\{f_i\}_0^n$ *satisfy conditions* (i)–(v) *and suppose that* $n \geqq 2$ *and* $p \neq 0$.

(a) *If* $n - p$ *is even then* $c_p \in R_0 = R^*$, i.e. *the unique* c_p*-optimal design is supported by the full set of Tchebycheff points* $s_0, s_1, \cdots, s_n$. *The* c_p*-optimal design and the minimal value of* $V(c_p, \xi)$ *are explicitly characterized in Theorem* 2.2.

(b) *For* $n - p$ *odd the unique* c_p*-optimal design is supported by the full set of Tchebycheff points* $t_0, t_1, \cdots, t_{n-1}$ *and this design is obtained from Theorem* 2.2 *by* *considering the 1st n component of* c_p. (*The vector* c_p *is in fact, contained in the set* $R_0(n-1) = R^*(n-1)$ *defined analogous to* R_0 *using* $\{t_i\}_0^{n-1}$ *instead of* $\{s_i\}_0^n$.)

The above theorem relies on the following result. For the ordinary polynomial case the proof is given in Natanson (1964), p. 53. The proof for the more general systems follows essentially word for word and will therefore be omitted.

THEOREM 4.2. *If* $n - p$ *is even then of all polynomials* $u = \sum_{i=0}^n a_i f_i$ *in which the coefficient of* f_p *is equal to unity the polynomial* $W_n(x)/a_p^{(n)}$ *minimizes*

$$\sup_{-1 \leqq x \leqq 1} |u(x)|.$$

If $n - p$ *is odd then the polynomial* $W_{n-1}(x)/a_p^{(n-1)}$ *has the above property.* (*The quantities* $a_p^{(n)}$ *and* $a_p^{(n-1)}$ *are the coefficients of* f_p *in* W_n *and* W_{n-1} *respectively*).

With the aid of the above theorem we can prove

LEMMA 4.1. *For* $n \geqq 2$ *the* c_p*-optimal design is supported by a subset of the Tchebycheff points* $s_0, s_1, \cdots, s_n$ *if* $n - p$ *is even and by a subset of the Tchebycheff points* $t_0, t_1, \cdots, t_{n-1}$ *if* $n - p$ *is odd.*

PROOF. By Remark 2.1 following Lemma 2.1 we have

$$\begin{aligned}
\inf_\xi V(c_p, \xi) &= \inf_\xi \sup_a (c_p, a)^2[\textstyle\int (a, f(x))^2 \xi(dx)]^{-1} \\
&= \sup_a \inf_\xi (c_p, a)^2[\textstyle\int (a, f(x))^2 \xi(dx)]^{-1} \\
&= \sup_a (c_p, a)^2[\sup_{-1 \leqq x \leqq 1} (a, f(x))^2]^{-1} \\
&= \begin{cases} (a_p^{(n)})^2 & n - p \quad \text{even} \\ (a_p^{(n-1)})^2 & n - p \quad \text{odd.} \end{cases}
\end{aligned}$$

Suppose that ξ_0 is c_p-optimal and $n - p$ is even. Then

$$\begin{aligned}
V(c_p, \xi_0) &= \sup_a (c_p, a)^2[\textstyle\int (a, f(x))^2 \xi_0(dx)]^{-1} \\
&\geqq (a_p^{(n)})^2[\textstyle\int (W_n(x))^2 \xi_0(dx)]^{-1} \\
&\geqq (a_p^{(n)})^2.
\end{aligned}$$

Since $|W_n(x)| = 1$ only for $s_0, s_1, \cdots, s_n$ strict inequality holds at the last step unless ξ_0 is supported by $s_0, s_1, \cdots, s_n$. The argument for $n - p$ odd is the same.

PROOF OF THEOREM 4.1. Suppose that $n - p$ is even, $n \geqq 2$ and $p \neq 0$. Since any c_p-optimal design is supported by $s_0, s_1, \cdots, s_n$ there must exist a solution $\{\epsilon_\nu p_\nu\}$ to the system of equations

$$\beta c_p = \textstyle\sum_{\nu=0}^n \epsilon_\nu p_\nu f(s_\nu), \qquad \beta^{-1} = |a_p^{(n)}|. \tag{4.1}$$

It suffices to show that $p_\nu \neq 0$ since in this case our assumption (ν) together with the Kiefer-Wolfowitz result (iv) described in Section 3 tells us that either the ϵ_ν alternate in sign or they are constant. However since $p \neq 0$ the first component of (4.1) reveals that ϵ_ν cannot be all of one sign. Therefore $c_p \in R_0 = R^*$.

Now suppose that $p_i = 0$ for some fixed i. From equation (4.1) the determinant with column vectors $f(s_0), \cdots, f(s_{i-1}), c_p, f(s_{i+1}), \cdots, f(s_n)$ is zero. Therefore there exists a polynomial $P(x) = \sum_0^n a_i f_i(x)$ such that $\sum_0^n a_i^2 > 0$, $a_p = 0$ and $P(s_\nu) = 0$ for $\nu \neq i$. Since $n - p$ is even, two cases may arise (a) n and p are both even or (b) n and p are odd. If n and p are even and $n = 2m$ we note, with the aid of condition (iii), that

$$Q(x) = P(x) + P(-x) = \sum_{k=0}^m a_{2k} f_{2k}(x), \qquad a_p = 0.$$

The polynomial $Q(x)$ has at most m terms since $a_p = 0$ and hence has at most $m - 1$ zeros on $(0, 1]$ by condition (iv). Note that $s_m = 0$ and $s_i = -s_{2m-i}$, $i = 0, 1, \cdots, m - 1$. If $i = m$, $Q(x)$ vanishes at $s_{m+1}, \cdots, s_{2m}$ implying that $a_{2k} = 0, k = 0, 1, \cdots, m$. If $i \neq m$ we suppose that $i > m$. In this case $Q(s_m) = 0$ and Q now has $m - 1$ terms and vanishes at $x = s_j, j = m + 1, \cdots, 2m, j \neq i$ again implying that $a_{2k} = 0$, $k = 0, 1, \cdots, m$. Therefore

$$P(x) = \sum_{k=1}^m a_{2k-1} f_{2k-1}$$

vanishes at $s_0, s_1, \cdots, s_{m-1}$ and hence at $s_{m+1}, \cdots, s_{2m}$. We conclude that $P(x) \equiv 0$ which is a contradiction. The case (b) where n and p are odd is handled in a similar manner using $Q(x) = P(x) - P(-x)$. The proof is omitted.

Now when $n - p$ is odd the c_p-optimal design is supported by a subset of $t_0, t_1, \cdots, t_{n-1}$ and hence there exists a solution $\{\epsilon_\nu p_\nu\}$ of

$$\beta c_p = \sum_{\nu=0}^{n-1} \epsilon_\nu p_\nu f(t_\nu), \qquad \beta^{-1} = |a_p^{(n-1)}|. \tag{4.2}$$

Omitting the last component the resulting system of equations reduces to the case $n - p$ even so that $p_\nu \neq 0$ and as before the ϵ_ν alternate in sign.

For the ordinary polynomials the explicit values of $\inf_\xi V(c_p, \xi)$ can be readily obtained from the coefficients of $W_n(x)$ or $W_{n-1}(x)$. Thus if $n - p$ is even and $k = 0, 1, \cdots, [\frac{1}{2}n]$ then

$$\inf_\xi V(c_{n-2k}, \xi) = \{n(n-k)^{-1}\tbinom{n-k}{k}2^{n-2k-1}\}^2$$

while for $n - p$ odd and $k = 0, 1, \cdots, [\frac{1}{2}(n-1)]$ we have

$$\inf_\xi V(c_{n-1-2k}, \xi) = \{(n-1)(n-1-k)^{-1}\tbinom{n-1-k}{k}2^{n-2k-2}\}^2.$$

We observe that for fixed k the minimal variance for estimating θ_k is equal when $n = k + 2i$ and $n = k + 2i + 1$, $i = 0, 1, \cdots$.

5. Remarks. We observe that the vectors $\sum_{\nu=0}^n \epsilon_\nu p_\nu f(s_\nu) = \beta c$ with ϵ_ν alternating or ϵ_ν of one sign form two sets of opposing n-dimensional convex faces in the boundary of $\mathfrak{R}$. Now each of these faces determines a convex cone such that any vector c in the cone has a c-optimal design supported on the Tchebycheff points. The simple property we wish to utilize is that a cone is closed under linear combinations with nonnegative coefficients. Each of these faces is, of course, the convex hull of the extreme points which are known; however it is of interest to determine, for example, in which face each of the vector c_p lie. For this purpose we consider the two cones

$$C_n = \{c \mid \beta c = \sum_{\nu=0}^n (-1)^{n-\nu} p_\nu f(s_\nu), \beta > 0\}$$

and

$$C_{n-1} = \{c \mid \beta c = \textstyle\sum_{\nu=0}^{n-1} (-1)^{n-1-\nu} q_\nu f(t_\nu), \beta > 0\}$$

where p_ν, $q_\nu \geqq 0$ and $\sum_\nu p_\nu = 1$ and $\sum_\nu q_\nu = 1$.

Suppose that $\{f_i\}_0^n$ is a T-system on $[-1, \infty)$ such that

$$F\begin{pmatrix} 0, 1, \cdots, n \\ x_0, x_1, \cdots, x_n \end{pmatrix} > 0 \tag{5.1}$$

whenever $-1 \leqq x_0 < x_1 < \cdots < x_n$. In this case $f(x) \varepsilon C_n$ for $x \geqq 1$ and hence $\sum_{i=1}^k \lambda_i f(x_i) \varepsilon C_n$ for $x_i \geqq 1$ and $\lambda_i \geqq 0$. In fact

$$\int f(x)\, d\mu(x) \varepsilon C_n \tag{5.2}$$

for any nonnegative measure μ on $[1, \infty)$ for which the integral is defined. The optimal designs for such linear combinations can easily be obtained using Theorem 2.2. For example the optimal design for the vector in (5.2) places mass

$$\alpha_\nu \left(\textstyle\sum_{j=0}^n \alpha_j\right)^{-1} \quad \text{at} \quad s_\nu \quad \text{where} \quad \alpha_\nu = \int |L_\nu(x)|\, d\mu(x), \qquad \nu = 0, 1, \cdots, n.$$

In order to determine in which face each of the vectors c_p lies we shall assume that for $k = n - 2, n - 1$, and n

$$F\begin{pmatrix} 0, 1, \cdots, k \\ x_0, x_1, \cdots, x_k \end{pmatrix} > 0 \tag{5.3}$$

whenever $-1 \leqq x_0 < \cdots < x_k \leqq 1$.

THEOREM 5.1. *Let $\{f_i\}_0^n$ satisfy conditions* (i)–(v) *of Section* 4 *and suppose that* (5.3) *holds. If $n - p$ is even then*

$$(-1)^k c_{n-2k} \varepsilon C_n, \qquad k = 0, 1, \cdots, [\tfrac{1}{2}n], \tag{5.4}$$

and if $n - p$ is odd then

$$(-1)^k c_{n-1-2k} \varepsilon C_{n-1}, \quad k = 0, 1, \cdots, [\tfrac{1}{2}(n - 1)].$$

Note that we have normalized the system $\{f_i\}_0^{n-1}$ in such a way that c_n is always in C_n. If $n = 4$ then c_4, $-c_2$, c_0 lie in C_n so that $\lambda_4 c_4 - \lambda_2 c_2 + \lambda_0 c_0 \varepsilon C_n$ when λ_0, λ_2, λ_4 are nonnegative and hence $\lambda_4\theta_4 - \lambda_2\theta_2 + \lambda_0\theta_0$ is optimally estimated by a design on the points $s_0, \cdots, s_4$.

PROOF OF THEOREM 5.1. The arguments are somewhat similar to that given in Theorem 4.1. We shall consider only the case where $n - p$ is even and $p \neq 0$. The remaining cases are similar.

Observe that $c_p \varepsilon C_n$ when $D_n(c_p) > 0$ and $-c_p \varepsilon C_n$ if $D_n(c_p) < 0$. Now the sign of $D_n(c_p)$ is the same as the sign of the coefficient of f_p in the polynomial

$$P(x) = F\begin{pmatrix} 0, 1, \cdots, n-1 & n \\ s_0, s_1, \cdots, s_{n-1} & x \end{pmatrix} = \textstyle\sum_{i=0}^n a_i f_i(x).$$

For $n = 2m$ the polynomial

$$Q(x) = 2^{-1}(P(x) + P(-x)) = \sum_{k=0}^{m} a_{2k}f_{2k}(x)$$

vanishes at $s_m = 0$ (so that $a_0 = 0$) and at the $m - 1$ points $s_{m+1}, \cdots, s_{2m-1}$.

Now the system $\{f_{2k}\}_1^m$ satisfies Descartes rule of signs (see Theorem 4.4, p. 25 in Karlin and Studden (1966a)). That is, the number of zeros of $Q(x)$ on $(0, 1]$ is bounded above by the number of sign changes in the sequence $a_2, a_4, \cdots, a_{2m}$ where zero terms are omitted if there are any. Since $Q(x)$ has $m - 1$ zeros the sequence $a_2, a_4, \cdots, a_{2m}$ must alternate in sign. Since $a_{2m} > 0$ we have $(-1)^k a_{n-2k} > 0, k = 0, \cdots, m$. Therefore $(-1)^k c_{n-2k} \in C_n$, for $k = 0, 1, \cdots, m$.

6. Counterexamples. It is natural to inquire about the necessity of the assumptions (i)–(v) imposed on our system of functions $\{f_0, f_1, \cdots, f_n\}$ for the validity of Theorem 4.1. The assumptions (iii) and (iv) are of particular interest. Concerning the necessity of something like assumption (iv), the following example was provided by Professor J. Kiefer. For $n = 3$ consider the estimation of θ_1 when $\{f_0, f_1, f_2, f_3\} = \{1, x, x^2, x^3 - x/2\}$ which satisfies all of the assumptions except (iv). It is not hard to prove that the unique c_1-optimum design is not supported on a subset of $\{s_j\}$ or $\{t_j\}$ and in fact is given by $\xi(\pm 2^{-\frac{1}{2}}) = \frac{1}{2}$. This follows from a direct computation which shows that the best Tchebycheff approximation of x on $[-1, 1]$ by a linear combination of 1, x^2 and $x^3 - x/2$ is $(x^3 - x/2)$, and the residual $x - (x^3 - x/2)$ attains its maximum in absolute value at $\pm 2^{-\frac{1}{2}}$.

The remainder of this section is devoted to an example concerning minimax designs. In the case where $f_i(x) = x^i, i = 0, 1, \cdots, n$, on $[-1, 1]$ it is known that the minimax design, which minimizes $\max_{-1 \leqq x \leqq 1} V(x, \xi)$, is supported by $n + 1$ points, namely the zeros of $(1 - x^2)P_n'(x) = 0$ where P_n is the nth Legendre polynomial. In fact in the ordinary polynomial case all admissible designs (see Kiefer (1959)) use only $n - 1$ points in the open interval $(-1, 1)$. The purpose of the following example is to exhibit a T-system for which any minimax design must be supported by *more* than $n + 1$ points. We shall utilize the theorem of Kiefer and Wolfowitz (1960) which states that ξ_0 minimizes $\max_{-1 \leqq x \leqq 1} V(x, \xi)$ if and only if ξ_0 maximizes the determinant of $M(\xi)$. Moreover $V(x, \xi_0) \leqq n + 1$ for all $x \in [-1, 1]$.

Consider the functions f_0, f_1, f_2 on $[-1, 1]$ where $f_0 \equiv 1$, $f_1(x) = x$ and f_2 is such that $f_2(x) \geqq 0$, $f_2(0) = 0$, $f_2(\pm 1) = 1$ and f_2 is convex. With these assumptions the design which minimizes $\max_{-1 \leqq x \leqq 1} V(x, \xi)$, among those concentrating all mass on 3 points, has equal mass at the points $-1, 0, 1$. This is geometrically clear after noting that for a design ξ with weights $\lambda_0, \lambda_1, \lambda_2$ on x_0, x_1, x_2 we have

$$|M(\xi)| = \lambda_0\lambda_1\lambda_2[\det |f_i(x_j)|_{i,j=0}^{2}]^2$$

and that the above determinant is the volume of the parallelopipe spanned by $f(x_0), f(x_1), f(x_2)$.

Now if the design ξ_0 with equal mass on $-1, 0, 1$ is minimax then

$$(f(x), M^{-1}(\xi_0)f(x)) \leqq 3 \quad \text{for} \quad x \in [-1, 1]. \tag{6.1}$$

However the quadratic surface

$$(1, y, z)M^{-1}(\xi_0)(1, y, z)' = 3$$

is given by

$$z = \tfrac{1}{3}\{2 \pm (4 - 3y^2)^{\frac{1}{2}}\}$$

which is strictly positive for $y \neq 0$. Thus (6.1) cannot hold when $f_2(x)$ is "sufficiently close" to zero. It can easily be verified that the system $\{1, x, x,^4\}$ does not satisfy (6.1) and hence the minimax design is on at least four points. Further analysis in this particular case shows that four points actually suffice.

REFERENCES

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 255–262.

HOEL, P. G. and LEVINE, A. (1964). Optimal spacing and weighing in polynomial prediction. *Ann. Math. Statist.* **35** 1553–1560.

KARLIN, S. and STUDDEN, W. J. (1966a). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Interscience, New York.

KARLIN, S. and STUDDEN, W. J. (1966b). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815.

KIEFER, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B.* **21** 273–319.

KIEFER, J. and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **12** 363–366.

KIEFER, J. and WOLFOWITZ, J. (1965). On a theorem of Hoel and Levine on extrapolation. *Ann. Math. Statist.* **36** 1627–1655.

NATANSON, I. P. (1964). *Constructive Function Theory* **1.** Ungar, New York.