The Annals of Statistics
1976, Vol. 4, No. 6, 1113–1123

OPTIMAL DESIGNS FOR LARGE DEGREE POLYNOMIAL REGRESSION

BY J. KIEFER[1] AND W. J. STUDDEN[2]

Cornell University and Purdue University

Polynomial regression of degree n on an interval is considered. Optimal designs ξ_n are discussed for various optimality criteria. The behavior of ξ_n for large n is investigated and comparisons of ξ_n with the limiting design ξ_0 are made.

1. Introduction. Let $f' = (f_0, f_1, \cdots, f_n)$ be a vector of linearly independent functions on a space $\mathscr{X}$. For each x or "level" in $\mathscr{X}$ an experiment can be performed whose outcome is a random variable $Y(x)$ with mean value $\beta' f(x) = \sum \beta_i f_i(x)$ and variance σ^2, independent of x. The functions f_i, $i = 0, 1, \cdots, n$ are called the regression functions and are assumed known to the experimenter while the vector of parameters $\beta' = (\beta_0, \beta_1, \cdots, \beta_n)$ and σ^2 are unknown. An experimental design is a probability measure ξ on $\mathscr{X}$. If ξ concentrates mass ξ_i at the points x_i, $i = 1, 2, \cdots, r$ and $\xi_i N = n_i$ are integers, the experimenter takes N uncorrelated observations, n_i at each x_i, $i = 1, 2, \cdots, r$. The covariance matrix of the least squares estimates of the parameters β_i is then given by $(\sigma^2/N)M^{-1}(\xi)$ where $M(\xi) = (m_{ij}(\xi))$, $m_{ij}(\xi) = \int f_i(x) f_j(x)\, d\xi(x)$ is the information matrix of the experiment or design ξ.

In experimental situations it may be desirable to use a design ξ which minimizes a particular functional of the matrix $M(\xi)$. For instance we may consider

(i) $|M^{-1}(\xi)|$ = determinant of $M^{-1}(\xi)$;

(ii) $\sup_{x \in \mathscr{X}} f'(x) M^{-1}(\xi) f(x)$; here $f'(x)M^{-1}(\xi)f(x)$ is proportional to the variance of the estimate of the regression function at x;

(iii) $f'(x_0)M^{-1}(\xi)f(x_0)$, or $c'M^{-1}(\xi)c$ for any vector c, or more generally $\operatorname{tr} CM^{-1}(\xi)$ where C is symmetric and positive semidefinite.

All of these are used in regression theory and are explained more fully in Fedorov (1972). It is known, see Kiefer and Wolfowitz (1960), that the minimization problems in (i) and (ii) are equivalent. Other results of this type are given in Fedorov (1972) and Kiefer (1974).

In the present paper we wish to examine the case where $\mathscr{X}$ is the interval $[-1, 1]$ and $f'(x) = (1, x, \cdots, x^n)$, i.e., the situation where our regression function is polynomial $\sum_{i=0}^n \beta_i x^i$. A given minimization problem of the type described

Received June 1975; revised March 1976.

[1] This research was supported by the NSF Grant MPS72-04998 A02. Reproduction is permitted in whole or in part for any purposes of the United States Government.

[2] Research support from NSF Grant 33552x2 and MP575-08294.

AMS 1970 *subject classification.* Primary 62K05.

Key words and phrases. Optimal design, regression, limiting design, generalized variance, extrapolation.

above, e.g., minimization of $|M^{-1}(\xi)|$, then gives rise to a sequence of designs ξ_n, one for each degree n. That is, for a fixed degree we suppose ξ_n satisfies

$$\min_\xi |M^{-1}(\xi)| = |M^{-1}(\xi_n)| .$$

We wish to examine the designs ξ_n for large values of n, in particular the limiting design ξ_0, if it exists, and to compare, say,

$$|M^{-1}(\xi_n)| \quad \text{and} \quad |M^{-1}(\xi_0)| \qquad \text{or}$$
$$\text{perhaps} \quad f'(x)M^{-1}(\xi_n)f(x) \quad \text{and} \quad f'(x)M^{-1}(\xi_0)f(x) .$$

The paper is motivated by an attempt to find single designs, for various criteria, which would work reasonably well for all degrees. This search led naturally to investigating whether there was any regularity in ξ_n for large n. In practice the regression function will usually be of low degree. It turns out, as in various other investigations, that the limiting design ξ_0 performs reasonably well even when used in a linear regression. For example the D-efficiency, using the generalized variance and defined below after Lemma 2.2, is about 70% and increases to the value one with increasing degree. Other, seemingly natural, definitions of efficiency involving $\sup_x f'(x)M^{-1}(\xi_0)f(x)$ are decreasing with the degree. We have indicated in Theorem 2.2 how the limiting efficiency in this case can be improved. The material here should also provide information on what to look for in more complex situations (see for example, Kiefer (1975) where quadratic regression on an n-simplex is considered).

In some situations sequential experimentation is costly but a large number of observations can be taken nonsequentially. If the degree is not known, we can obtain some idea of the cost, from using a design which is optimum for a very large degree, over what one could achieve with an approximately optimum sequential design which would determine the right degree with high probability and act accordingly. The material should also be of some theoretical interest.

The paper is divided into three more sections. In Section 2 we consider the determinant of the information matrix $|M(\xi)|$ as mentioned above. The extrapolation problem minimizing $f'(x_0)M^{-1}(\xi)f(x_0)$ for a fixed x_0 outside of $\mathscr{X}$ is considered in Section 3. Results similar to those given in Section 2 and 3 for the optimal designs for estimating the separate coefficients β_i are discussed in Section 4.

2. The generalized variance. Consider the case where

$$f'(x) = (1, x, \cdots, x^n) \equiv \bar{f}_n'(x) \quad (\text{say}) , \qquad \mathscr{X} = [-1, 1] ,$$
$$M_n(\xi) = \textstyle\int_{-1}^{1} f(x)f'(x)\, d\xi(x)$$

and maximize the determinant $|M_n(\xi)|$. The design ξ_n maximizing $|M_n(\xi)|$ is called D-optimal. The following theorem is given in Fedorov (1972), page 91. We include a short proof here for completeness.

THEOREM 2.1. *The sequence ξ_n, $n = 1, 2, \cdots$ of D-optimal designs converges weakly to ξ_0 where ξ_0 has density $1/\pi(1 - x^2)^{\frac{1}{2}}$.*

PROOF. The proof follows fairly readily from a number of known results. The D-optimal design ξ_n concentrates mass $(n+1)^{-1}$ on the $n+1$ zeros x_v, $v = 0, 1, \cdots, n$ of $(1-x^2)P_n'(x)$, where P_n is the nth Legendre polynomial. See Karlin and Studden (1966a). If θ_v are defined by $x_v = \cos\theta_v$, $0 \leqq \theta_v \leqq \pi$, it is known (see Erdös and Turán (1940)) that the θ_v become uniformly distributed on the half circle $0 \leqq \theta \leqq \pi$ in the sense that if $N(a, b)$ denotes the number of θ_v in $[c, d]$ then

$$n^{-1}N(c, d) \rightarrow \frac{|d-c|}{\pi}.$$

It then follows that $\xi_n \rightarrow \xi_0$ where ξ_0 is the distribution of $Y = \cos X$ and X is uniform on $(0, \pi)$. The statement of the theorem then follows.

Let $d_n(x, \xi) = \bar{f}_n'(x)M_n^{-1}(\xi)\bar{f}_n(x)$ and denote $\sup_x d_n(x, \xi)$ by $d_n(\xi)$. The quantity $d_n(x, \xi)$ is proportional to the variance of our estimate of the regression curve at the point x assuming our regression was of degree n and we used the design ξ. We will compare $d_n(x, \xi_0)$ and $d_n(x, \xi_n)$ for small values of n and consider whether ξ_0 is "asymptotically optimal" in some sense. We also compare $|M_n(\xi_n)|$ and $|M_n(\xi_0)|$.

A general calculation for $d_n(x, \xi_0)$ can be made. The quantity $d_n(x, \xi_0)$ is invariant under a change of basis for our functions $1, x, x^2, \cdots, x^n$. We use as a basis the polynomials which are orthonormal on $[-1, 1]$ with the weight or measure ξ_0. These are the polynomials $1, 2^{\frac{1}{2}}T(x), \cdots, 2^{\frac{1}{2}}T_n(x)$ where $T_k(\cos\theta) = \cos k\theta$ are the Chebyshev polynomials of the 1st kind. In this case we find that

$$\begin{aligned} d_n(x, \xi_0) &= 1 + 2\textstyle\sum_{k=1}^{n} T_k^2(x) \\ &= n + \tfrac{1}{2} + \frac{1}{2}\frac{\sin(2n+1)\theta}{\sin\theta} \qquad (2.1) \\ &= n + \tfrac{1}{2} + \tfrac{1}{2}U_{2n}(x) \end{aligned}$$

where $U_k(\cos\theta) = \sin(k+1)\theta/\sin\theta$ are the Chebyshev polynomials of the 2nd kind.

It is known that $d_n(\xi_n) = \sup_x d_n(x, \xi_n) = n + 1$ and that the sup is attained at the points where ξ_n concentrates its equal mass. These are the zeros of $(1-x^2)P_n'(x)$, where P_n denotes the Legendre polynomial. If we use ξ_0 instead of ξ_n we have that $d_n(\xi_0) = n + \frac{1}{2} + \frac{1}{2}\sup_x U_{2n}(x)$. Since $\sup_x U_k(x) \leqq k + 1$ (see Davis (1963)) it follows that $d_n(\xi_0) = 2n + 1$. This value is about double the value $n + 1$. The sup here is reached only at the end points $x = \pm 1$.

We shall consider both functions $d_n(x, \xi_n)$ and $d_n(x, \xi_0)$ for $n = 1, 2$ and 3. For $n = 1$ the D-optimal design has equal mass at $x = \pm 1$. Simple calculations show that

$$d_1(x, \xi_1) = 1 + x^2 \qquad \text{and} \qquad d_1(x, \xi_0) = 1 + 2x^2$$

so that $d_1(x, \xi_1) \leqq d_1(x, \xi_0)$ for all x. The next case, $n = 2$, seems to be somewhat more indicative of the general situation. Here the D-optimal design has mass $\frac{1}{3}$

on points $x = 1, 0, 1$. Calculations then give

$$d_2(x, \xi_2) = 3 - \tfrac{9}{2}x^2 + \tfrac{9}{2}x^4$$
$$d_2(x, \xi_0) = 3 - 6x^2 + 8x^4$$

and we have $d_2(x, \xi_0) \leqq d_2(x, \xi_2)$ for $|x| \leqq (\frac{3}{7})^{\frac{1}{2}} = .655$. Thus the approximate D-optimal design is better for x in the middle of our interval $[-1, 1]$. For $n = 3$ the situation is more complicated, as expected. The D-optimal design has equal weight $\frac{1}{4}$ on ± 1 and $\pm 1/5^{\frac{1}{2}} = \pm .447$. More calculations give $d_3(x, \xi_3) = 3.248 + 8.261x^2 - 26.267x^4 + 18.756x^6$ while $d_3(x, \xi_0) = \frac{7}{2} + \frac{1}{2}(\sin 7\theta/\sin\theta)$, $x = \cos\theta$. The values are roughly comparable in the range $|x| \leqq 0.9$. A small table of values is given below. Both functions are symmetric about zero.

θ	90	80	70	60	50	40	30	20	10	0
$x = \cos$	0	.174	.342	.5	.643	.766	.866	.94	.98	1
$d(x, \xi_3)$	3.25	3.47	3.88	3.96	3.50	2.84	2.58	2.98	3.66	3
$d_3(x, \xi_0)$	3	3.33	3.91	4.00	3.39	2.73	3.00	4.44	6.20	7

If we define the G-efficiency (see Atwood 1969) of a design ξ as $(n + 1)/d_n(\xi)$ we then note that the limiting design ξ_0 has G-efficiency $(n + 1)/(2n + 1)$ which, unexpectedly, decreases to the value $\frac{1}{2}$. It is natural to inquire whether there is a design with limiting G-efficiency equal to one. In this regard we have

THEOREM 2.2. *For each* $\varepsilon > 0$ *there exists a design* ξ_ε *such that*

$$\liminf_{n\to\infty} \frac{n+1}{d_n(\xi_\varepsilon)} \geqq 1 - \varepsilon\,. \tag{2.2}$$

PROOF. The result is obtained using the following lemma from Szego (1959), page 31.

LEMMA 2.1. *Let* $\rho(x)$ *be a polynomial of degree* l *on* $[-1, 1]$ *and write* $\rho(\cos\theta) = |h(e^{i\theta})|^2$ *where* $x = \cos\theta$, $h(z)$ *is of degree* l, $h(z) \neq 0$ *for* $|z| < 1$ *and* $h(0) > 0$. *Let* $h(e^{i\theta}) = c(\theta) + is(\theta)$ *and* $w(x) = (d/\pi)(1 - x^2)^{-\frac{1}{2}}/\rho(x)$ *where* d *is such that* $\int w(x)\,dx = 1$. *Then the polynomials orthonormal with respect to* $w(x)$ *are given by* $p_0(x) = 1$ *and*

$$p_k(\cos\theta) = \left(\frac{2}{d}\right)^{\frac{1}{2}} \{c(\theta)\cos k\theta + s(\theta)\sin k\theta\} \qquad k \geqq 1\,.$$

The idea is to change the measure ξ_0 by putting some mass near ± 1. This is accomplished by changing ξ_0 to a design ξ_ε with density $w(x)$ where $\rho(x)$ depends on $\delta(\varepsilon)$. For fixed $\delta = \delta(\varepsilon)$ we let $\rho(x)$ be a polynomial such that

$$\begin{aligned} \delta^{\frac{1}{2}} &\leqq \rho(x) \leqq 1 + \delta && \text{for} \quad -1 \leqq x \leqq 1\,, \\ 1 &\leqq \rho(x) \leqq 1 + \delta && \text{for} \quad -1 + \delta \leqq x \leqq 1 - \delta\,, \\ \delta^{\frac{1}{2}} &\leqq \rho(x) \leqq 2\delta^{\frac{1}{2}} && \text{for} \quad 1 - |x| < \delta/2\,. \end{aligned}$$

The proof of the theorem will be to show that $(n + 1)/d_n(\xi_\varepsilon) \geqq 1 - \varepsilon$ for large

n when $\delta = \delta(\varepsilon)$ is taken sufficiently small. Such a polynomial exists by the Weierstrass theorem and we denote its degree by l. Applying the above lemma we then have

$$d_n(x, \xi_\varepsilon) - 1 = \sum_1^n p_k^2(\cos\theta)$$
$$= \frac{2}{d}\{c^2(\theta)\sum_1^n \cos^2 k\theta + s^2(\theta)\sum_1^n \sin^2 k\theta + c(\theta)s(\theta)\sum_1^n \sin 2k\theta\}\,.$$

Inserting the values $\cos^2 k\theta = (1 + \cos 2k\theta)/2$ and $\sin^2 k\theta = (1 - \cos 2k\theta)/2$ and using $\rho(\theta) = c^2(\theta) + s^2(\theta)$ we obtain

$$\sum_1^n p_k^2(\cos\theta) = d^{-1}\{n\rho(\theta) + (c^2(\theta) - s^2(\theta))\sum_1^n \cos 2k\theta + 2c(\theta)s(\theta)\sum_1^n \sin 2k\theta\}\,.$$

From the choice of $\rho(\theta)$ it is easily seen that $d = d(\delta) \to 1$ as $\delta \to 0$ so that the first term is $d^{-1}n\rho(\theta) = n(1 + o(\delta))$. The other two terms can be readily handled. Since $\rho(\theta) = c^2(\theta) + s^2(\theta)$ it follows that $|c^2(\theta) - s^2(\theta)| \leqq 2\rho(\theta)$ and $|c(\theta)s(\theta)| \leqq \rho(\theta)$. The remaining two terms are then bounded by

$$2d^{-1}\rho(\theta)\{|\sum_1^n \cos 2k\theta| + |\sum_1^n \sin 2k\theta|\}\,.$$

Now for $1 - |x| < \delta/2$, $\rho(\theta)$ is small and the term in brackets is bounded by $2n$. For $1 - |x| > \delta/2$ the bracket can be bounded by a term depending on δ but not on n. Therefore the remaining two terms can be bounded by $f(\delta) + no(\delta)$. The result then follows.

We now turn to a comparison of the two determinants $|M_n(\xi_n)|$ and $|M_n(\xi_0)|$. The ratio can be calculated in a fairly explicit form. We let $D_n(\xi) = |M_n(\xi)|$.

THEOREM 2.3. *If ξ_n is the D-optimal design and ξ_0 has the* arcsin *density as described in Theorem* 2.1 *then*

$$\frac{D_{n+1}(\xi_{n+1})}{D_{n+1}(\xi_0)} = 2n^{\frac{1}{4}}e^{\delta_n} \tag{2.3}$$

where

$$\delta_n = \tfrac{1}{4}\left\{\sum_{k=1}^n \frac{1}{k} - \ln n\right\} - \sum_{k=2}^\infty (-1)^k \frac{\zeta_n(k)}{k(k+1)}\left(1 - \frac{1}{2^k}\right) \tag{2.4}$$

and

$$\zeta_n(k) = \sum_{l=1}^n \frac{1}{l^k}\,, \qquad \zeta(k) = \sum_{l=1}^\infty \frac{1}{l^k}\,.$$

Before proving Theorem 2.3 we shall state an additional relevant lemma and make a few additional remarks.

LEMMA 2.2 (Rubin). *If δ_n is defined above in* (2.4) *then*

$$\delta_n \to \delta = \tfrac{1}{2} - \tfrac{13}{12}\log 2 - 3\zeta'(-1)$$
$$\approx -.00464602\,.$$

The proof of Lemma 2.2 will be omitted. The quantity $\zeta'(-1)$ is the derivative of the zeta function at -1 and has value $\zeta'(-1) = -.165421142$. (See

Walther (1926)). If one wished to use the ratio (2.3), the only complicated quantity is δ_n. If we replace δ_n by δ and consider

$$\rho_{n+1} = \frac{D_{n+1}(\xi_0)}{2n^{\frac{1}{2}}e^{\delta}D_{n+1}(\xi_{n+1})}$$

then $\rho_2 \approx 1.00465$ and ρ_n seems to be decreasing to one. (The limit of course is equal to one.) Moreover the quantity $e^{\delta} \approx .99536$ so that the ratio in (2.3) is essentially $2n^{\frac{1}{2}}$.

In terms of efficiency, the appropriate quantity is the D-efficiency of ξ_0 (see Atwood (1969)) defined by

$$E_n = \left(\frac{D_n(\xi_0)}{D_n(\xi_n)}\right)^{1/(n+1)} .$$

The values of E_n for $n = 1, 2, 3, 4$ can be easily calculated to be 0.71, 0.75, 0.79 and 0.81 respectively. By Theorem 2.3 it is easy to see that ξ_0 has limiting D-efficiency equal to one. This immediately raises the question about the relationship between the G-efficiency and the D-efficiency of a design. We note that by the Kiefer–Wolfowitz equivalence theorem D-optimality or D-efficiency equal to one is equivalent to G-optimality or G-efficiency equal to one. Using inequalities (see Kiefer, 1960) of the type

$$\frac{D_n(\xi)}{D_n(\xi_n)} \geqq \exp\{n + 1 - d_n(\xi)\}$$

one can show that a limiting G-efficiency of one produces a limiting D-efficiency of one. The converse however is not true as our example shows. (A limiting G-efficiency of $1 - \varepsilon$ produces a limiting D-efficiency of $\exp -\varepsilon/(1 - \varepsilon)$.)

We note finally that the design ξ_ε (used in Theorem 2.2 to give a better G-efficiency than the limiting design ξ_0) should presumably have a better D-efficiency than ξ_0 calculated from equation (2.3). This however is not the case as will be seen from Theorem 2.4.

PROOF OF THEOREM 2.3. From Szego (1959), page 28, we find the value $D_n(\xi_0) = \prod_{r=1}^{n} k_r^{-2}$ where $k_n = 2^{\frac{1}{2}}2^{n-1}$ is the coefficient of x^n in $2^{\frac{1}{2}}T_n(x)$: these being the polynomials orthonormal to ξ_0. In this case $D_n(\xi_0) = 2^{-n^2}$. The value $D_n(\xi_n)$ is given by (see Karlin and Studden (1966b))

$$D_n(\xi_n) = 2^{n(n+1)}(\textstyle\prod_{v=1}^{n} v^v)^4 n^{-n} \prod_{v=1}^{2n} v^{-v} .$$

Letting

$$R_n = \frac{D_n(\xi_n)}{D_n(\xi_0)} ,$$

a straightforward calculation gives

$$\frac{R_{n+1}}{R_n} = \frac{(1 + 1/n)^{n+1}}{(1 + 1/2n)^{2n+1}} .$$

Then

$$\log \frac{R_{l+1}}{R_l} = (l+1)\log\left(1+\frac{1}{l}\right) - (2l+1)\log\left(1+\frac{1}{2l}\right)$$
$$= \frac{1}{4l} - \sum_{k=2}^{\infty} \frac{(-1)^k}{l^k k(k+1)}\left(1-\frac{1}{2^k}\right),$$

and

$$\log \frac{R_{n+1}}{R_1} = \sum_{l=1}^{n} \log \frac{R_{l+1}}{R_l}$$
$$= \frac{1}{4}\sum_{l=1}^{n}\frac{1}{l} - \sum_{k=2}^{\infty}\frac{\zeta_n(k)(-1)^k}{k(k+1)}\left(1-\frac{1}{2^k}\right).$$

Since $R_1 = 2$ it then follows that

$$\log \frac{R_{n+1}}{2n^{\frac{1}{4}}} = \frac{1}{4}\left\{\sum_{l=1}^{n}\frac{1}{l} - \ln n\right\} - \sum_{k=2}^{\infty}\frac{\zeta_n(k)(-1)^k}{k(k+1)}\left(1-\frac{1}{2^k}\right),$$

and Theorem 2.3 follows.

The asymptotic behavior of $D_n(\xi)$ for designs ξ with densities can be ascertained from results on Hankel determinants. The next result follows nearly immediately from Grenander and Szego (page 84) and Szego (page 142).

THEOREM 2.4. *If ξ has density $g(x)$ then*

$$D_n^{1/(n+1)}(\xi) \approx D_n^{1/(n+1)}(1)e^{G(g)},$$

where

$$G(g) = \int_{-1}^{1} \frac{\log g(x)}{\pi(1-x^2)^{\frac{1}{2}}}\,dx,$$

and

$$D_n(1) = 2^{-n(n-1)}\prod_{\nu=1}^{n}\nu^{3\nu-2n}(n+\nu)^{n-\nu}.$$

The approximation is taken in the sense that the ratio of both sides tends to one.

A variational argument shows that $G(g)$ is maximized by ξ_0 or $g_0(x) = 1/\pi(1-x^2)^{\frac{1}{2}}$. It then follows that g_0 is the only density which has asymptotic D-efficiency equal to one. Thus the designs ξ_ε used in Theorem 2.2 have asymptotic D-efficiency less than one.

3. Extrapolation. In this section we consider the minimization of

$$d_n(x_0,\xi) = \bar{f}_n'(x_0)M_n^{-1}(\xi)\bar{f}_n(x_0), \qquad |x_0| > 1. \tag{3.1}$$

Using the design ξ and assuming an nth degree regression, the quantity $d_n(x_0, \xi)$ is proportional to the variance of the least squares estimate of the regression at the point x_0. Since $|x_0| > 1$ and observations are confined to $[-1, 1]$ we have an extrapolation problem. It is known, see Hoel and Levine (1964) or Studden (1968), that the optimal design ξ_n minimizing $d_n(x_0, \xi)$ concentrates mass p_v on the points $s_v = \cos v\pi/n$, $v = 0, 1, \cdots, n$. The value p_v is given by

$$p_v = \frac{|l_v(x_0)|}{\sum_{v=0}^{n}|l_v(x_0)|}$$

where $l_v(x)$, $v = 0, 1, \cdots, n$ are the Lagrange polynomials of degree n specified by $l_v(s_\mu) = \delta_{\mu v}$, i.e., fixing v, $l_v(x)$ vanishes at all the values s_l except s_v where it has the value one.

THEOREM 3.1. *If ξ_n is the optimal extrapolation design then $\xi_n \to \xi_0$ where ξ_0 has density*

$$\frac{(x_0^2 - 1)^{\frac{1}{2}}}{\pi|x_0 - x|(1 - x^2)^{\frac{1}{2}}}.$$

PROOF. We consider only the case $x_0 > 1$. The measure ξ_n has weights proportional to $|l_v(x_0)|$ at $s_v = \cos v\pi/n$. The Lagrange polynomial l_v is given by

$$l_v(x_0) = \frac{\prod_{\mu=0}^n (x_0 - s_\mu)}{(x_0 - s_v) \prod_{\mu \neq v} (s_\mu - s_v)}.$$

Since the numerator is constant we see that ξ_n is proportional to a measure

$$\frac{1}{x_0 - x} d\rho_n(x)$$

where $d\rho_n$ has mass $[\prod_{\mu \neq v} (s_\mu - s_v)]^{-1} = \gamma_v$ at $s_v = \cos v\pi/n$. Substituting the values $s_v = \cos v\pi/n$ in γ_n and using the law

$$\cos A - \cos B = -2 \sin (A + B)/2 \sin (A - B)/2$$

a straightforward calculation reveals that the values $\gamma_0, \gamma_1, \cdots, \gamma_n$ are proportional to $1, 2, 2, \cdots, 2, 1$. It then follows as in Theorem 2.1 that the limiting measure has density proportional to

$$\frac{1}{(x_0 - x)(1 - x^2)^{\frac{1}{2}}}. \tag{3.2}$$

The CRC tables (*Handbook for Probability and Statistics*, 2nd edition, page 609, formula 215) give a value of $\pi/(x_0^2 - 1)^{\frac{1}{2}}$ for the integral of (3.2) over the range -1 to 1. The theorem then follows.

We will now make a comparison of $d_n(x_0, \xi_n)$ and $d_n(x_0, \xi_0)$. From Studden (1968) we know that $d_n(x_0, \xi_n) = T_n^2(x_0)$ where $T_n(x)$ is the Chebyshev polynomial of the first kind. In order to evaluate $d_n(x_0, \xi_0)$ we use the fact as noted above that the expression for it given in (3.1) is invariant under a basis change so that

$$d_n(x_0, \xi_0) = \sum_{k=0}^n p_k^2(x_0) \tag{3.3}$$

if we let $p_k(x)$ denote the polynomials orthonormal with respect to the measure ξ_0.

In order to evaluate (3.3) we use Lemma 2.1. We apply this result with $\rho(x) = (x_0 - x)$. We find that $c(\theta) = (a + b \cos \theta)$, $s(\theta) = b \sin \theta$ where a and b satisfy the two conditions

$$a^2 + b^2 = x_0, \qquad 2ab = -1. \tag{3.4}$$

The restrictions on $h(z)$ in the lemma give

$$a^2 = \tfrac{1}{2}(x_0 + (x_0^2 - 1)^{\frac{1}{2}}), \qquad b^2 = \tfrac{1}{2}(x_0 - (x_0^2 - 1)^{\frac{1}{2}}). \tag{3.5}$$

This then gives

$$p_k(\cos\theta) = \left(\frac{2}{c}\right)^{\frac{1}{2}} \{(a + b\cos\theta)\cos k\theta + b\sin\theta\sin k\theta\} \tag{3.6}$$
$$= \left(\frac{2}{c}\right)^{\frac{1}{2}} \{a\cos k\theta + b\cos(k-1)\theta\}$$

where $c = (x_0^2 - 1)^{\frac{1}{2}}$.

In order to evaluate $\sum_{k=0}^{n} p_k^2(x_0)$ we use (3.6), some half-angle trigonometric formulae, some series summation from Jolley (1961), and simplify the following expression:

$$\begin{aligned}
&\textstyle\sum_{k=1}^{n} (a\cos k\theta + b\cos(k-1)\theta)^2 \\
&\quad = \textstyle\sum_{1}^{n} [a^2\cos^2 k\theta + b^2\cos^2(k-1)\theta + 2ab\cos k\theta\cos(k-1)\theta] \\
&\quad = b^2 + a^2\cos^2 n\theta + x_0 \textstyle\sum_{k=1}^{n-1}\cos^2 k\theta - \sum_{k=1}^{n}\cos k\theta\cos(k-1)\theta \\
&\quad = b^2 + a^2\cos^2 n\theta + x_0\left(\frac{n-1}{2} + \frac{\cos n\theta\sin(n-1)\theta}{2\sin\theta}\right) \\
&\qquad - \tfrac{1}{2}\textstyle\sum_{k=1}^{n}(\cos(2k-1)\theta + \cos\theta) \\
&\quad = b^2 + a^2\cos^2 n\theta + x_0\left(\frac{n-1}{2} + \frac{\cos n\theta\sin(n-1)\theta}{2\sin\theta}\right) \\
&\qquad - \tfrac{1}{2}\left(\frac{1}{2}\frac{\sin 2n\theta}{\sin\theta} + n\cos\theta\right) \\
&\quad = b^2 + a^2\cos^2 n\theta - \frac{x_0}{2} - \tfrac{1}{2}\cos n\theta\cos(n-1)\theta .
\end{aligned}$$

Using equation (3.5) and $T_n(\cos\theta) = \cos n\theta$ we then have

$$\textstyle\sum_{k=0}^{n} p_k^2(x_0) = \frac{1}{(x_0^2-1)^{\frac{1}{2}}}[(x_0 + (x_0^2-1)^{\frac{1}{2}})T_n(x_0) - T_n(x_0)T_{n-1}(x_0)] .$$

The following theorem can then be readily deduced.

THEOREM 3.2. *Let $d_n(x_0, \xi)$ be as in* (3.1) *and let $r_n(x_0) = d_n(x_0, \xi_n)/d_n(x_0, \xi_0)$. Then*

$$r_n^{-1}(x_0) = \frac{1}{(x_0^2-1)^{\frac{1}{2}}}\left\{x_0 + (x_0^2-1)^{\frac{1}{2}} - \frac{T_{n-1}(x_0)}{T_n(x_0)}\right\} .$$

For linear regression we have $r_1^{-1}(x_0) = 1 + (x_0^2 - 1)^{\frac{1}{2}}/x_0$; while for quadratic regression $r_2^{-1}(x_0) = 1 + 2x_0(x_0^2 - 1)^{\frac{1}{2}}/(2x_0^2 - 1)$. Note as expected that both of these values are near 1 for x_0 close to 1. Bounds and other limit relations can be obtained using the following two lemmas.

LEMMA 3.1. *The ratio $a_n = T_n(x_0)/T_{n-1}(x_0)$ is increasing to $x_0 + (x_0^2 - 1)^{\frac{1}{2}}$.*

PROOF. Since $T_{n+1}(x_0) = 2x_0T_n(x_0) - T_{n-1}(x_0)$ we have $a_{n+1} = 2x_0 - 1/a_n$ and then a_n is seen to be increasing using induction. The actual limit value follows from the fact that (see Szego (1959), page 189)

$$\lim_{n\to\infty} \frac{(x_0 + (x_0^2-1)^{\frac{1}{2}})^n}{2T_n(x_0)} = 1 .$$

LEMMA 3.2. *For fixed x_0, the ratio $r_n(x_0)$ decreases to the value $\frac{1}{2}$.*

The proof follows immediately from Lemma 3.1. Note in particular that $\frac{1}{2} \leqq r_n(x_0)$ and of course $r_n(x_0) \leqq 1$. One can show further that $\lim_{x_0 \to 1} r_n(x_0) = 1$ and $\lim_{x_0 \to \infty} r_n(x_0) = \frac{1}{2}$.

The asymptotic behavior of $d_n(x_0, \xi)$ can also be evaluated for any design ξ with density using results from Szego (1959).

THEOREM 3.3. *If the design ξ has density g then*

$$d_n(x_0, \xi) \approx (2\pi)^{-1} D_g^{-2}(z_0^{-1}) \frac{z_0^{2n} - 1}{z_0^{2n}},$$

where $z_0 = x_0 + (x_0^2 - 1)^{\frac{1}{2}}$, and

$$D_g(z) = \exp \left\{ \frac{1}{4\pi} \int_{-\pi}^{\pi} \log [g(\cos \theta)|\sin \theta|] \frac{1 + ze^{-i\theta}}{1 + ze^{-i\theta}} d\theta \right\}.$$

The approximation means the ratio converges to one.

PROOF. From Szego (1959), page 295 we find that if $p_n(x_0)$ denotes the nth polynomial orthonormal with respect to g then

$$p_n(x_0) \approx (2\pi)^{-\frac{1}{2}} z^n \{D(z^{-1})\}^{-1}.$$

The result then follows by a simple Abelian argument since

$$d_n(x_0, \xi) = \sum_{k=0}^{n} p_k^2(x_0).$$

COROLLARY. *Among designs which are absolutely continuous with respect to Lebesgue measure, the one which asymptotically minimizes $d_n(x_0, \xi)$ has density*

$$g_0(x) = \frac{(x_0^2 - 1)^{\frac{1}{2}}}{\pi |x_0 - x| (1 - x^2)^{\frac{1}{2}}}.$$

PROOF. This result follows by noting that $D_g(z_0)$ must be real. We therefore maximize $|D_g(z_0)|^2$. A variational argument shows this to be $g_0(x)$.

4. Individual coefficients. Analyses similar to Sections 2 and 3 can be given for the estimation of the individual coefficients. Since a more complete investigation of the asymptotic properties of the information matrix is presently being made, we shall be somewhat briefer and proofs will be omitted.

We consider the estimation of β_k in the model $\sum_{k=0}^{n} \beta_k x^k$. The variance of the LSE using a design ξ is denoted by $V(n, k, \xi)$ and the optimal design is devoted by $\xi(n, k)$.

It can be shown that for k fixed, $\xi(n, n - k)$ has limit density $g_0(x) = \pi^{-1}(1 - x^2)^{-\frac{1}{2}}$ while if $k/n = q$ and q is fixed with $0 < q < 1$ then $\xi(n, k)$ has limiting density proportional to $[q^2 + (1 - q^2)x^2]^{-1}(1 - x^2)^{-\frac{1}{2}}$. In the case $k = n$, let $\xi_n = \xi(n, n)$ and ξ_0 denote the design with density $g_0(x)$. Then

$$\frac{V(n, n, \xi_n)}{V(n, n, \xi_0)} \to \frac{\pi}{2} D^2(0) \tag{4.1}$$

where $D(z)$ is defined in Section 3. The value $D(0)$ is maximized by the design with density g_0. The efficiency given in (4.1) is constant and equal to $\frac{1}{2}$. This value "agrees" with the remark in Section 3 that $r_n(\xi_0) \to \frac{1}{2}$ (the case $x_0 \to \infty$ and the highest coefficient are equivalent problems in certain respects).

For k fixed the designs $\xi(n, k)$ degenerate to having all mass at the origin. Thus for estimating the slope at the origin or the coefficient β_1 the optimal design concentrates its mass closer and closer to zero as $n \to \infty$. The actual rate of this convergence has been ascertained.

Acknowledgments. We wish to thank Professor Herman Rubin for proving Lemma 2.2 and for other helpful conversations. We also thank Professor Szego for providing basic ideas and the referee and editors for their helpful comments.

REFERENCES

[1] DAVIS, PHILIPS J. (1963). *Interpolation and Approximation.* Blaisdell, Waltham, Mass.

[2] ERDÖS, P. and TURÁN, P. (1940). Interpolation III. *Ann. of Math.* **41** 510–553.

[3] FEDOROV, V. V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

[4] GRENANDER, ULF and SZEGO, GABOR (1958). *Toeplitz Forms and Their Applications.* Univ. of California Press.

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

[6] JOLLEY, L. B. W. (1961). *Summation of Series.* Dover, New York.

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

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

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

[10] KIEFER, J. (1966). Optimum experimental designs V. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 381–405, Univ. of California Press.

[11] KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Math. Statist.* **2** 849–879.

[12] KIEFER, J. (1975). Optimal design: Variation in structure and performance under change of criterion. *Biometrika* **62** 277–288.

[13] STUDDEN, W. J. (1968). Optimal designs on Tchebycheff points. *Ann. Math. Statist.* **39** 1435–1447.

[14] SZEGO, S. (1959). *Orthogonal Polynomials.* Amer. Math. Soc. Colloq. Publ. **23** New York.

[15] WALTHER, ALWIN (1926). Anschauliches zur Riemannschen Zeta funktion. *Acta Math.* **48** 393–400.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14850

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
LAFAYETTE, INDIANA 47907