The Annals of Statistics
1993, Vol. 21, No. 2, 767–771

A NOTE ON *E*-OPTIMAL DESIGNS FOR WEIGHTED POLYNOMIAL REGRESSION

BY HOLGER DETTE

Universität Göttingen

In a recent paper Pukelsheim and Studden determined the E-optimal design for the polynomial regression model on the interval $[-1,1]$ where the variances of different observations are assumed to be constant. In this note we show that these results can be generalized for polynomial regression models with non constant variances proportional to specific functions.

1. Introduction. Consider the polynomial regression model of degree $d \in \mathbb{N}$

$$g(x) = \theta' f(x), \qquad x \in [-1,1],$$

where $f(x) = (1, x, \dots, x^d)'$ is the vector of regression functions and $\theta = (\theta_0, \dots, \theta_d)'$, is the vector of unknown parameters. A design ξ is a probability measure on $[-1,1]$ (or on its sigma field). ξ is called an exact design if ξ puts masses (n_i/n) $(\sum_{i=1}^{s} n_i = n)$ at the points $x_i \in [-1,1]$, $i = 1, \dots, s$. In this case the experimenter takes n_i uncorrelated observations at the point x_i with expectation $g(x_i)$ and variance $\sigma^2/\lambda(x_i)$, $i = 1, \dots, s$. The function $\lambda\colon [-1,1] \to [0,\infty)$ is called the efficiency function [see Fedorov (1972), page 66] and in this note we will assume that λ is one of the functions 1, $1+x$, $1-x$ and $1-x^2$. The information matrix of a design ξ is defined by

$$M_d(\xi) = \int_{-1}^{1} f(x) f'(x) \lambda(x)\, d\xi(x) \in \mathbb{R}^{(d+1)\times(d+1)},$$

where

$$\lambda(x) = (1+x)^u (1-x)^v \qquad u, v \in \{0,1\}. \tag{1.1}$$

For an exact design the covariance matrix of the least squares estimate of the unknown parameter vector θ is proportional to $M_d^{-1}(\xi)$ and an optimal design maximizes (or minimizes) a concave (or convex) function depending on $M_d(\xi)$ (or its inverse).

This paper deals with the E-optimality criterion for the parameter vector θ. More precisely, a design ξ is called E-optimal for θ if ξ maximizes the minimum eigenvalue of the information matrix $M_d(\xi)$. In a recent paper Pukelsheim and Studden (1993) proved that for a constant efficiency function (i.e., $u = v = 0$) the E-optimal design for θ is supported at the points $s_j = \cos(\pi(d-j)/d)$, $j = 0, \dots, d$. In this note we will show that similar results

Received February 1992; revised September 1992.
AMS 1991 *subject classification*. Primary 62K05.
Key words and phrases. Weighted polynomial regression E-optimal designs, Chebyshev polynomials.

hold also for the efficiency functions defined by (1.1). These results allow us to present a simple example in which the support points of the E-optimal design cannot be transformed linearly when the design space is changed from the interval $[-1, 1]$ to $[-b, b]$ where $b > 1$ is sufficiently large.

2. *E*-optimal designs for weighted polynomial regression. For optimality criteria depending on the determinants of the information matrix $M_d(\xi)$ most of the results for the constant efficiency function carry over to efficiency functions of the form (1.1) [e.g., see Studden (1982) or Dette (1990)]. This is a consequence of the structure of the supporting hyperplanes to the moment space $\mathscr{M}_{2d} = \{(c_0, \dots, c_{2d})' | c_i = \int_{-1}^{1} x^i \, d\xi(x), \ i = 0, \dots, 2d\}$ [see Karlin and Shapely (1953)]. Especially the results for D_1-optimality carry over to the efficiency function (1.1). Because the D_1-optimal design for a constant efficiency function is supported at the same set of points s_j as the E-optimal design, it is reasonable to investigate if the support of the D_1-optimal design also coincides with the support of the E-optimal design for the nonconstant efficiency function in (1.1).

Pukelsheim and Studden (1993) showed that the E-optimal design for polynomial regression with a constant efficiency function is supported at the zeros $s_j = \cos(\pi((d-j)/d))$, $j = 0, \dots, d$, of the polynomial $(1-x^2)U_{d-1}(x)$ where

$$U_{d-1}(x) = \frac{\sin(d \arccos x)}{\sin(\arccos x)}, \qquad x \in [-1, 1]$$

denotes the Chebyshev polynomial of the second kind (with leading coefficient 2^{d-1}) orthogonal with respect to the measure $\sqrt{1-x^2}\, dx$ [see Szegö (1975), page 60 or Chihara (1978), page 1]. An essential step in their proof is to show that the Chebyshev polynomial of the first kind $T_d(x) = \cos(d \arccos x)$ defines the supporting hyperplane at the inball vector of the Elfving set $\mathscr{R} = co(\{\varepsilon f(x) | x \in [-1, 1], \ \varepsilon = \pm 1\})$ [here $co(A)$ denotes the convex hull of the set A]. For the efficiency functions $\lambda(x) = 1 - x^2$, $1 - x$ and $1 + x$ the analogues of the Chebyshev polynomial of the first kind are the polynomials $U_d(x)$,

$$\frac{d}{\sqrt{2}} B\left(\frac{1}{2}, d\right) P_d^{(1/2, -1/2)}(x) = \frac{\sin(((2d+1)/2)\Theta)}{\sqrt{2}\sin(\Theta/2)},$$

$$\frac{d}{\sqrt{2}} B\left(\frac{1}{2}, d\right) P_d^{(-1/2, 1/2)}(x) = \frac{\cos(((2d+1)/2)\Theta)}{\sqrt{2}\cos(\Theta/2)},$$

where $\Theta = \arccos x$, $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ denotes the beta-function and $P_n^{(\alpha, \beta)}(x)$ the nth Jacobi polynomial orthogonal with respect to the measure $(1-x)^\alpha(1+x)^\beta$ [see Szegö (1975), page 60]. The proof of the following theorem is performed by similar arguments as in the case of a constant efficiency function and therefore omitted.

THEOREM 2.1. *Let* $\lambda(x) = (1+x)^u(1-x)^v$, $u, v \in \{0, 1\}$, *and* $b = (b_0, \dots, b_d)'$ *denote the vector of the coefficients of the polynomials*

$$(2.1)\quad 2^{-\lfloor (u+v)/2 \rfloor}\left(d + \left\lfloor \frac{u+v}{2} \right\rfloor\right) B\left(\frac{1}{2}, d + \left\lfloor \frac{u+v}{2} \right\rfloor\right) P_d^{(-1/2+v,\,-1/2+u)}(x).$$

The E-optimal design for θ *is unique and supported at the zero's* $s_j = \cos(\pi(2d - 2j + v)/(2d + u + v))$, $j = 0, \dots, d$, *of the polynomial*

$$(1+x)^{1-u}(1-x)^{1-v} P_{d-1+u+v}^{(1/2-v,\,1/2-u)}(x).$$

Moreover, the minimum eigenvalue of the E-optimal information matrix has multiplicity 1 *and is given by* $\|b\|^{-2}$. *The weights of the support points* s_j *are given by* $w_j = u_j/\|b\|^2$ *where the vector* $u = (u_0, \dots, u_d)'$ *is the unique solution of*

$$\sum_{j=0}^{d} (-1)^{d-j} u_j (1+s_j)^{u/2}(1-s_j)^{v/2} f(s_j) = b.$$

REMARK 2.2. For the efficiency function $\lambda(x) = 1 - x^2$ the proof of Theorem 2.1 can be performed exactly in the same way as the proof of the corresponding result ($\lambda(x) = 1$) given by Pukelsheim and Studden (1993) because this efficiency function preserves the symmetry properties of the problem. In the case of a nonsymmetric efficiency function some minor changes are necessary because the coefficients corresponding to the powers x^{d-j} in the polynomials $P_d^{(-1/2,1/2)}(x)$ and $P_d^{(1/2,-1/2)}(x)$ have different sign patterns compared to the symmetric case. Note that Pukelsheim and Studden (1993) originally stated and proved their result for E-optimal designs for parameter subsystems of the vector θ which satisfy some "regularity" condition. Theorem 2.1 can also be proved for special subsystems of θ but for the sake of brevity only the case of the full parameter vector θ is stated in this note.

REMARK 2.3. Observing the identities

$$T_d(x) = dB\left(\tfrac{1}{2}, d\right) P_d^{(-1/2,1/2)}(x),$$
$$U_d(x) = \frac{d+1}{2} B\left(\tfrac{1}{2}, d+1\right) P_d^{(1/2,1/2)}(x),$$

Theorem 2.1 generalizes the results of Pukelsheim and Studden (1993) to the nonconstant efficiency functions $\lambda(x) = 1 - x^2, 1 - x, 1 + x$. Note that Theorem 2.1 is *not* valid for efficiency functions in (1.1) with positive $u, v > 0$ although there exist generalizations in this direction for the D-optimality criterion. More precisely, it is well known that the D-optimal design for the efficiency function in (1.1), $u, v > 0$, puts equal masses at the zeros of the Jacobi polynomial $P_{d+1}^{(v-1,u-1)}(x)$ [see, e.g., Karlin and Studden (1966), page 339]. For $u, v \in (0, \infty) \setminus \{1\}$ analogous results for the D_1- and E-optimality criterion do not exist [see Dette (1992) for some numerical examples].

REMARK 2.4. It follows from the proof of Theorem 2.1 that the vector $b/\|b\|^2$ defines an inball vector of the corresponding Elfving set [see Elfving (1952)]

$$\mathscr{R}^{(u,v)} := co\left\{\varepsilon(1+x)^{u/2}(1-x)^{v/2}f(x)|x \in [-1,1], \varepsilon = \mp 1\right\}$$

which is the direction where the largest ball centered at the origin and included in $\mathscr{R}^{(u,v)}$ touches the boundary of $\mathscr{R}^{(u,v)}$. This yields to some nice extremal properties of the polynomials defined in (2.1) [see Pukelsheim and Studden (1993) for the case $u = v = 0$].

EXAMPLE 2.5. The proof of Theorem 2.1 can be transferred to all intervals of the form $[-b, b]$ where $b < 1$ but not to all intervals with $b > 1$. Moreover, the support points s_j^* of the E-optimal design on $[-b, b]$ are not proportional to the support points s_j of the optimal design on $[-1, 1]$ provided that b is sufficiently large. To see this, let $d = 1$ and $u = v = 1$, that is, $\lambda(x) = b^2 - x^2$. If $b = 1$ we obtain by Theorem 2.1 that the E-optimal design puts equal masses at the points $s_0 = -1/\sqrt{2}$ and $s_1 = 1/\sqrt{2}$. For an arbitrary $b > 0$ we have still that the E-optimal design is symmetric and supported at two points, say $-a$ and a, $a > 0$. The information matrix of this design is given by

$$M(\xi) = \begin{pmatrix} b^2 - a^2 & 0 \\ 0 & a^2(b^2 - a^2) \end{pmatrix}.$$

It follows by elementary calculations that for $b \leq \sqrt{2}$ the optimal design puts equal masses at the points $s_0^* = -b/\sqrt{2}$ and $s_1^* = b/\sqrt{2}$ and that the minimum eigenvalue is $b^4/4$ with multiplicity 1 (the points s_0^* and s_1^* are proportional to s_0 and s_1). If $b > \sqrt{2}$ we obtain that the E-optimal design on $[-b, b]$ has equal masses at the points $s_0^* = -1$ and $s_1^* = 1$ (independent of the length of the interval) and the minimum eigenvalue is given by $b^2 - 1$ with multiplicity 2. In this case the support points of the E-optimal design are not obtained by a linear transformation from the corresponding points on the interval $[-1, 1]$. It is remarkable that the threshold $b = \sqrt{2}$ also applies for the quadratic regression on $[-b, b]$ with a constant efficiency $\lambda(x) \equiv 1$ [see Galil and Kiefer (1977), (3.8), page 34]. Recently, Heiligers (1992) showed that the threshold for the cubic model with a constant efficiency is approximately 1.61918 while it follows by some algebra that the thresholds for the quadratic and cubic model with efficiency function $\lambda(x) = b^2 - x^2$ are given by $\sqrt{2}$ and approximately 1.59003, respectively.

Acknowledgments. The author would like to thank the referees and an Associate Editor for their careful reading of the manuscript. Their constructive comments yielded a considerable improvement in the presentation of the results.

REFERENCES

CHIHARA, T. S. (1978). *An Introduction to Orthogonal Polynomials*. Gordon and Breach, New York.

DETTE, H. (1990). A generalization of D- and D_1-optimal design in polynomial regression. *Ann. Statist.* **18** 1784–1804.

DETTE, H. (1992). Experimental design for a class of weighted polynomial regression models. *Comput. Statist. Data Anal.* **14** 359–373.

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

FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. Academic, New York.

GALIL, J. and KIEFER, J. (1977). Comparision of rotatable designs on balls, I (Quadratic). *J. Statist. Plann. Inference* **1** 27–40.

HEILIGERS, B. (1992). E-optimal polynomial regression designs. Habilitationsschrift, RWTH, Aachen.

KARLIN, S. and SHAPELY, L. S. (1953). *Geometry of Moment Spaces*. *Mem. Amer. Math. Soc.* **12**. Amer. Math. Soc., Providence, RI.

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

PUKELSHEIM, F. and STUDDEN, W. J. (1993). E-optimal designs for polynomial regression. *Ann. Statist.* **21** 402–415.

STUDDEN, W. J. (1982). Optimal designs for weighted polynomial regression using canonical moments. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **2** 335–350. Academic, New York.

SZEGÖ, G. (1975). *Orthogonal Polynomials* **23**, *4th ed*. Amer. Math. Soc., Providence, RI.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT GÖTTINGEN
LOTZESTRASSE 13
W-3400 GÖTTINGEN
GERMANY