The Annals of Statistics
1989, Vol. 17, No. 2, 618–623

NOTE ON SOME ϕ_p-OPTIMAL DESIGNS FOR POLYNOMIAL REGRESSION[1]

BY W. J. STUDDEN

Purdue University

In a recent paper of Gaffke an example was given regarding the ϕ_p-optimal designs for the highest two coefficients in a one dimensional polynomial regression. The purpose of this paper is to supply a direct proof of this result using the theory of canonical moments and orthogonal polynomials.

1. Introduction. Consider a simple polynomial regression model on $[-1,1]$. Thus for each $x \in [-1,1]$ an observation may be observed with mean value $\sum_{i=0}^{m}\theta_i x^i$ and constant variance σ^2, independent of x. The parameters $\theta' = (\theta_0, \dots, \theta_m)$ and σ^2 are unknown. An experimental design is a probability measure ξ on $[-1,1]$. If N uncorrelated observations are taken and ξ has mass $\xi(i) = n_i N^{-1}$ at x_i, $i = 1, \dots, r$, then n_i observations are taken at x_i. The covariance matrix of the least squares estimates of θ is given by $(\sigma^2/N)M^{-1}(\xi)$ where $M(\xi)$ is the information matrix of the design ξ given by

$$m_{ij} = \int_{-1}^{1} x^{i+j}\, d\xi(x). \tag{1.1}$$

Generally speaking the design ξ is chosen to "maximize" $M(\xi)$ or "minimize" $M^{-1}(\xi)$. Amongst criteria for this minimization are Kiefer's ϕ_p-criteria [see Kiefer (1974), (4.18) or Kiefer (1975), page 337]. The function ϕ_p is the "p-mean" of $M^{-1}(\xi)$ given by

$$\begin{aligned} \phi_p(M) &= \left\{(m+1)^{-1}\operatorname{tr} M^{-p}(\xi)\right\}^{1/p} \\ &= \left\{(m+1)^{-1}\left(\sum_{v=0}^{m}\lambda_v^{-p}\right)\right\}^{1/p}, \end{aligned} \tag{1.2}$$

where the λ_v are the eigenvalues of $M(\xi)$ and $-1 < p \le \infty$.

If one is only interested in a subset of the parameters, say the highest s parameters, then we write

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

where M_{22} is $s \times s$. The information matrix regarding these parameters is given

Received October 1987; revised April 1988.
[1]Research supported by NSF Grant DMS-85-03771.
AMS 1980 *subject classifications*. Primary 62K05; secondary 49B40.
Key words and phrases. Approximate design theory, Tchebycheff approximation, canonical moments.

by

$$\Lambda(\xi) = \Lambda = M_{22} - M_{21}M_{11}^{-1}M_{12}. \tag{1.3}$$

The corresponding ϕ_p-criterion is to minimize

$$\phi_p(\Lambda) = \{s^{-1}\operatorname{tr}\Lambda^{-p}\}^{1/p}. \tag{1.4}$$

The values $p = 0$, 1 and ∞ are usually singled out. These values correspond to $(\det \Lambda^{-1})^{1/s}$, $\operatorname{tr}(\Lambda^{-1})/s$ and the maximum eigenvalue of Λ^{-1}, respectively. Since all the optimal designs we encounter have $m + 1$ support points for which M is nonsingular, we shall not discuss nonsingular M here. See Gaffke (1987) or Pukelsheim (1980) for details and further references on this question.

The present paper is concerned with the case $s = 2$ for polynomial regression. The ϕ_p-optimal designs were given rather explicitly in a recent paper of Gaffke (1987) which considers general regression models and is concerned with "the characterizations of design optimality and admissibility for partial parameter estimation." The special case $m = 2$ was considered by Pukelsheim (1980). The result in question is stated in the following theorem. The polynomials $T_k(x)$ and $U_k(x)$ denote the usual Tchebycheff polynomials of the first and second kind; see, e.g., Abramowitz and Stegun (1964).

THEOREM 1 (Gaffke). *The ϕ_p-optimal design for the highest two coefficients in polynomial regression of degree m on $[-1,1]$ concentrates mass at the $m + 1$ zeros $x_0 = -1 < x_1 < \cdots < x_{m-1} < x_m = 1$ of*

$$(1 - x^2)(U_{m-1}(x) + \beta U_{m-3}(x)), \tag{1.5}$$

where U_{-1} is defined to be zero and β is the root of

$$\left(\frac{1-\beta}{2}\right)^{p+1} - \beta = 0, \qquad 0 \le \beta < 1. \tag{1.6}$$

The corresponding weights are given by

$$\xi(x_j) = (1 - \beta^2)/\left\{(m-1)(1-\beta^2) + (1+\beta)^2 - 4\beta T_{m-1}^2(x_j)\right\} \tag{1.7}$$

for $j = 1, \dots, m-1$ and

$$\xi(-1) = \xi(+1) = \tfrac{1}{2}(1-\beta^2)/\left\{(m-1)(1-\beta^2) + (1-\beta)^2\right\}. \tag{1.8}$$

The proof of Theorem 1 as given in Gaffke (1987) is rather elaborate and ingenious and is an application of more general results concerning partial parameter estimation. The purpose of this paper is to give a more direct proof. The proof deals directly with the moments or rather the canonical moments of the design ξ. The theory of canonical moments allows us to "identify" the ϕ_p-optimal design rather quickly. The identification or equivalence with the form in Theorem 1 is then "straightforward" but somewhat algebraically involved. For the theory of canonical moments the reader is referred to Lau (1983); see also Lau and Studden (1985), Studden (1980, 1982) and Skibinsky (1968).

2. Proof of Theorem 1. In order to prove the theorem a short description of the canonical moments and a statement of some of the results is needed. For an arbitrary design ξ the information matrix $M(\xi)$, and hence also $\Lambda(\xi)$, depends on the moments

$$c_i = \int_{-1}^{1} x^i \, d\xi(x), \qquad i = 1, 2, \dots, 2m.$$

The canonical moments are defined as follows. For a given set of moments $c_0, c_1, \dots, c_{i-1}$ let c_i^+ denote the maximum of the ith moment $\int x^i \, d\mu(x)$ over the set of all probability measures μ having moments $c_1, c_2, \dots, c_{i-1}$. Similarly let c_i^- denote the corresponding minimum. The canonical moments are defined by

$$p_i = \frac{c_i - c_i^-}{c_i^+ - c_i^-}, \qquad i = 1, 2, \dots . \tag{2.1}$$

Note that $0 \le p_i \le 1$. We will have $0 < p_i < 1$ whenever $c_1, c_2, \dots, c_i$ is in the interior of the corresponding moment space. Whenever $p_i = 0$ or 1 the subsequent p_k, $k > i$, are left undefined. As an example consider the Jacobi measure with density proportional to $(1 + x)^\alpha (1 - x)^\beta$ ($\alpha > -1$, $\beta > -1$). For this measure

$$p_{2k} = \frac{k}{\alpha + \beta + 2k + 1}, \qquad k > 0,$$

$$p_{2k+1} = \frac{\alpha + k + 1}{\alpha + \beta + 2k + 2}, \qquad k > 0.$$

The uniform measure ($\alpha = \beta = 0$) has $p_{2k+1} = \frac{1}{2}$, $k \ge 0$ and $p_{2k} = k/(2k + 1)$. The arc-sin distribution ($\alpha = \beta = -\frac{1}{2}$) has $p_k = \frac{1}{2}$ for all k.

Since the underlying interval is $[-1, 1]$ and $\phi_p(\Lambda)$ is convex in M and invariant under reflexion of the design we may assume that any ϕ_p-optimal design is symmetric. In this case all the odd moments of ξ are zero and $\Lambda(\xi)$ reduces to

$$\begin{aligned} \Lambda &= M_{22} - M_{21} M_{11}^{-1} M_{12} \\ &= \begin{pmatrix} a_{m-1} & 0 \\ 0 & a_m \end{pmatrix}, \end{aligned} \tag{2.2}$$

where $a_k = \int P_k^2(x) \, d\xi(x)$, $k = m - 1, m$ and $\{P_i\}$ is the sequence of polynomials, with leading coefficient 1, which are orthogonal to $d\xi(x)$. The fact that $\Lambda(\xi)$ is diagonal follows from elementary matrix calculations. The formula for the two components a_k, $k = m - 1, m$ is the $L_2(\xi)$ squared norm of x^k minus its projection onto the linear space spanned by $1, x, \dots, x^{m-2}$. For $k = m - 1$ this is the defining property of P_{m-1} and hence $a_{m-1} = \int P_{m-1}^2 \, d\xi$. Since the odd moments of ξ are zero the projection of x^m does not "involve" x^{m-1} so that $a_m = \int P_m^2 \, d\xi$. In terms of the canonical moments a_{m-1} and a_m are given (for symmetric ξ) by

(2.3) $$\int_{-1}^{1} P_k^2(x)\, d\xi(x) = \prod_{i+1}^{k} p_{2i} q_{2(i-1)},$$

where $q_i = 1 - p_i$, $1 \le i \le k$, and $q_0 = 1$. Equation (2.3) follows from Theorem 2.4.8 of Lau (1983) where we use the fact that if the measure ξ has odd moments zero then $p_{2i+1} = \frac{1}{2}$ whenever it is defined.

To obtain the ϕ_p = optimal design in terms of the p_i we may now minimize

$$\left(a_{m-1}^{-p} + a_m^{-p}\right)^{1/p}$$

with respect to p_i. This leads immediately to the following lemma. For $p = \infty$ we are simply maximizing a_m since clearly $a_{m-1} > a_m$.

Lemma 2.1. *The ϕ_p-optimal design ξ_p is given by $p_{2m} = 1$, $p_i = \frac{1}{2}$, $i = 1, 2, \dots, 2m - 1$, $i \neq 2m - 2$ and $p_{2m-2} = (1 + \beta)/2$ where β satisfies* (1.6).

Lemma 2.1 gives, in a sense, a complete solution to the ϕ_p-optimal design problem in the present situation. It is, however, relatively straightforward to go from the form given in Lemma 2.1 in terms of the canonical moments to the support and weights of the design ξ_p given by Gaffke in Theorem 1. The remainder of the proof is a brief description of procedure.

In the case that $p_{2m} = 1$ it is known that the corresponding measure has support at ± 1 and $m - 1$ points on the interor $(-1, 1)$ [see, e.g., Karlin and Studden (1966), Chapter 4]. The $m - 1$ interior points are the roots of the polynomial Q_{m-1} where $\{Q_k\}$ is the sequence of polynomials orthogonal to $(1 - x^2)\, d\xi_p$. If ξ_p is symmetric, then $p_{2i+1} = \frac{1}{2}$ for all i and these polynomials (with leading coefficient equal to 1) are defined recursively by

(2.4) $$Q_{k+1}(x) = xQ_k(x) - p_{2k} q_{2k+2} Q_{k-1}, \qquad k \ge 1,$$

where $Q_0 \equiv 1$ and $q_i = 1 - p_i$ [see Lau (1983), Remark 2.4.4, page 31].

We now note (as remarked earlier) that the sequence $p_i = \frac{1}{2}$ for all $i \ge 1$ corresponds to the arcsin measure

$$d\mu_0 = \frac{dx}{\pi\sqrt{1 - x^2}}.$$

The corresponding orthogonal polynomials are the Tchebycheff polynomials $T_k(x)$ of the first kind. The polynomials orthogonal to $(1 - x^2)\, d\mu_0$ correspond to the Tchebycheff polynomials of the second kind denoted by $U_k(x)$. [$U_k(x)$ has leading coefficient 2^k.] Since ξ_p has canonical moments $p_i = \frac{1}{2}$ for $i \le 2m - 3$, it follows that, for $i \le m - 2$, $U_i(x) = 2^i Q_i(x)$. Inserting $k = m - 2$ in (2.4) we find that

(2.5) $$2^{m-1} Q_{m-1}(x) = U_{m-1}(x) + \beta U_{m-3}(x).$$

Thus the support of ξ_p is on the zeros to $(1 - x^2)[U_{m-1}(x) + \beta U_{m-3}(x)]$ as stated in Theorem 1.

The remaining question concerns the weights $\xi_p(x_j)$ given in Theorem 1. For the interior points we use the fact that

$$\xi_p^{-1}(x_j) = \left(1 - x_j^2\right) \sum_{k=0}^{m-2} \left(Q_k^*(x_j)\right)^2, \tag{2.6}$$

where Q_k^* are ortho*normal* with respect to $(1 - x^2)\, d\xi_p(x)$. The weight at ± 1 is given by

$$\xi_p(\pm 1) = 2 \sum_{k=0}^{m-1} \left(R_k^*(1)\right)^2, \tag{2.7}$$

where R_k^* are ortho*normal* with respect to $(1 + x)\, d\xi_p(x)$. Formulas (2.6) and (2.7) are given in Karlin and Studden [(1966), Chapter 4, pages 115 and 116]. Equation (2.7) is the last equation in Theorem 3.1, while (2.6) is the second to last formula on page 116. Both equations require transforming results from $[0, 1]$ onto $[-1, 1]$.

To convert (2.6) to (1.7) we use the fact that $\int U_k^2(x)(1 - x^2)\, d\mu_0 = \frac{1}{2}$. Then $Q_k^*(x) = \sqrt{2}\, U_k(x)$ for $k \le m - 3$. Note that the normalizing factor for Q_{m-2} uses $p_{2m-2} = 1 - q_{2m-2}$. It can be shown that

$$\left(Q_{m-2}^*\right)^2 = U_{m-2}^2(x)/q_{2m-2}.$$

Using a number of trigonometric formulas (2.6) will reduce to (1.7). The details of this reduction are omitted. The considerations of (2.7) are somewhat similar and are also omitted. □

Acknowledgments. Thanks go to the editors and referees for a very careful reading of this paper.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A., eds. (1964). *Handbook of Mathematical Functions. Appl. Math. Ser.* **55**. National Bureau of Standards, Washington.

GAFFKE, N. (1987). Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression. *Ann. Statist.* **15** 942–957.

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

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

KIEFER, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.) 333–354. North-Holland, Amsterdam.

LAU, T. S. (1983). Theory of canonical moments and its application in polynomial regression, I and II. Technical Reports 83-23 and 83-24, Dept. Statistics, Purdue Univ.

LAU, T. S. and STUDDEN, W. J. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments. *Ann. Statist.* **13** 383–394.

PUKELSHEIM, F. (1980). On linear regression designs which maximize information. *J. Statist. Plann. Inference* **4** 339–364.

SKIBINSKY, M. (1968). Extreme nth moments for distributions on $[0, 1]$ and the inverse of a moment space map. *J. Appl. Probab.* **5** 693–701.

STUDDEN, W. J. (1980). D_s-optimal designs for polynomial regression using continued fractions. *Ann. Statist.* **8** 1132–1141.

STUDDEN, W. J. (1982). Some robust type D-optimal designs in polynomial regression. *J. Amer. Statist. Assoc.* **77** 916–921.

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907