The Annals of Statistics
1994, Vol. 22, No. 2, 890–903

DISCRIMINATION DESIGNS FOR POLYNOMIAL REGRESSION ON COMPACT INTERVALS

BY HOLGER DETTE

Universität Göttingen

In the polynomial regression model of degree $m \in \mathbb{N}$ we consider the problem of determining a design for the identification of the correct degree of the underlying regression. We propose a new optimality criterion which minimizes a weighted p-mean of the variances of the least squares estimators for the coefficients of x^l in the polynomial regression models of degree $l = 1, \dots, m$. The theory of canonical moments is used to determine the optimal designs with respect to the proposed criterion. It is shown that the canonical moments of the optimal measure satisfy a (nonlinear) equation and that the support points are given by the zeros of an orthogonal polynomial. An explicit solution is given for the discrimination problem between polynomial regression models of degree $m-2$, $m-1$ and m and the results are used to calculate the discrimination designs in the sense of Atkinson and Cox for polynomial regression models of degree $1, \dots, m$.

1. Introduction. Consider the polynomial regression situation on the interval $[a, b]$. For each $x \in [a, b]$ an experiment can be performed whose outcome is a random variable $Y(x)$ with expectation

$$E\big(Y(x)\big) = \sum_{i=0}^{m} a_i x^i$$

and variance σ^2 independent of x. A design ξ is a probability measure on $[a, b]$, and the matrix

$$M_m(\xi) = \int_a^b (1, x, \dots, x^m)'(1, x, \dots, x^m)\, d\xi(x)$$

is the information matrix of ξ. If ξ concentrates masses ξ_i at the points x_i, $i = 1, \dots, r$, and $N\xi_i = n_i$ are integers, the experimenter takes N uncorrelated observations, n_i at each x_i, $i = 1, \dots, r$, and the covariance matrix of the least squares estimator for the parameters a_i is proportional to the inverse of the information matrix $M_m(\xi)$.

An optimal design maximizes or minimizes some functional depending on the information matrix or its inverse, and there are numerous criteria which can be used for determining a "good" design ξ [see, e.g., Kiefer (1974) or Silvey (1980)]. Optimal designs in polynomial regression have been studied in considerable detail [see, e.g., Hoel (1958), Guest (1958), Murty and Studden (1972) and Studden (1980, 1982a, b, 1989)]. All these articles assume that the

Received October 1992; revised March 1993.
AMS 1991 *subject classifications.* Primary 62K05; secondary 62J05.
Key words and phrases. Canonical moments, discrimination design, scalar-optimality.

degree of the polynomial is known before the experiments are carried out. It is the purpose of this paper to consider design problems when the degree of the underlying polynomial regression cannot be fixed in advance and a design has to be constructed for the discrimination between the rival models of degree $1, \dots, m$. Similar problems were extensively studied by Atkinson (1972), Atkinson and Cox (1974) and Atkinson and Fedorov (1975a, b). In Section 2 a new optimality criterion is introduced which can be used for determining discrimination designs for the polynomial regression models of degree $l = 1, \dots, m$. An optimal design with respect to this criterion minimizes a weighted p-mean of the variances of the least squares estimators for linear combinations of the parameter vectors in different models. In Section 3 we will use the proposed criterion for the calculation of optimal designs for model discrimination in the polynomial regression model. We present a complete solution of the proposed problem using the theory of canonical moments which was introduced (in the context of optimal design theory) by Studden (1980, 1982a, b) [see also Lau (1983, 1988)]. It is shown that the optimal designs for the discrimination between polynomials up to degree m is supported at the zeros of a certain set of orthogonal polynomials, and we obtain explicit representations of the optimal design for the discrimination between regression models of degree $m-2$, $m-1$ and m. Finally, some examples are presented in Section 4, and we determine the optimal discriminating designs in the sense of Atkinson and Cox (1974) for polynomial regression models of degree $l = 1, \dots, m$. It is also shown that these designs converge weakly to the arcsine distribution when the degree of the regression tends to infinity.

2. Generalized c-optimal designs. Let $f_l(x) = (1, x, \dots, x^l)'$ denote the vector of monomials up to degree $l \in \mathbb{N}$, and let $a_l = (a_{l0}, \dots, a_{ll})' \in \mathbb{R}^{l+1}$ denote the vector of unknown parameters for the polynomial regression of degree l

$$g_l(x) = a_l' f_l(x), \qquad l = 1, \dots, m.$$

The information matrix in the model g_l is now given by $M_l(\xi) = \int_a^b f_l(x) f_l'(x)\, d\xi(x)$ and an efficient design for all models should maximize or minimize a function depending on the information matrices $M_1(\xi), \dots, M_m(\xi)$ or their inverses. Let $c_l \in R^{l+1}$ be given vectors, and let β_l be nonnegative numbers with sum 1, $\beta_m > 0$, and $p < 1$. Then we say a design ξ is $\Phi^c_{p,\beta}$-optimal with respect to the prior β if and only if ξ maximizes the weighted p-mean

$$\Phi^c_{p,\beta}(\xi) := \left[\sum_{l=1}^{m} \beta_l \left(c_l' M_l^-(\xi) c_l \right)^{-p} \right]^{1/p} \tag{2.1}$$

and allows the estimability of all linear combinations $c_l' a_l$, $l = 1, \dots, m$. The prior $\beta = (\beta_1, \dots, \beta_m)$ reflects the experimenter's belief about the adequacy of the different models $g_1(x), \dots, g_m(x)$. The optimality criterion (2.1) was introduced by Dette (1993) in a more general context investigating geometric characterizations of the optimal designs with respect to the criterion (2.1). To obtain an optimal design for model discrimination, we choose the special vectors $b_l = (0, \dots, 0, 1)' \in \mathbb{R}^{l+1} (l = 1, \dots, m)$ as "the vectors for the highest coefficient" in the model g_l. Assume that the experimenter uses the following stepwise procedure for the selection of the variables in the model: starting with a linear regression, the F-test for a quadratic trend H_0: $a_{22} = 0$ is performed. If this test rejects H_0, then the additional parameter a_{22} is included into the model, and the hypothesis of a cubic trend H_0: $a_{33} = 0$ is tested. This procedure continues until the specified bound m for the maximum degree of the polynomial regression is reached or terminates if one of the tests accepts its corresponding hypothesis. Because an optimal design for $b_l' a_l = a_{ll}$ [i.e., the design that minimizes $b_l' M_l^-(\xi) b_l$] also maximizes the power of the F-test for H_0: $a_{ll} = 0$, a $\Phi_{p,\beta}^b$-optimal design with respect to the prior β will be useful for discriminating between different degrees of the polynomial regression [note that the noncentrality parameter of the F-test for H_0: $a_{ll} = 0$ is proportional to $a_{ll}^2 (b_l' M_l^-(\xi) b_l)^{-1}$]. In the following we will call such a design $\Phi_{p,\beta}^b$-optimal discriminating design with respect to the prior β.

Note that in the situation considered here our approach generalizes a criterion proposed by Atkinson and Cox (1974) [see also (Atkinson (1972)]. These authors considered an "extended" model and constructed optimal designs for detecting departures from the given models which have to be discriminated. In the situation described so far, the extended model is given by $g_m(x)$ while the rival models are $g_1(x), \dots, g_{m-1}(x)$. Atkinson and Cox (1974) proposed to maximize the weighted product

$$\prod_{l=1}^{m-1} \left(\frac{\det M_m(\xi)}{\det M_l(\xi)} \right)^{1/(m-l)} \tag{2.2}$$

Observing that $b_l' M_l^{-1}(\xi) b_l = \det M_{l-1}(\xi) / \det M_l(\xi)$ and that the case $p = 0$ in (2.1) has to be understood as the limit $p \to 0$, we obtain that a $\Phi_{0,\beta}^b$-optimal discriminating design maximizes

$$\Phi_{0,\beta}^b(\xi) := \lim_{p \to 0} \Phi_{p,\beta}^b(\xi) = \prod_{l=1}^{m} \left(\frac{\det M_l(\xi)}{\det M_{l-1}(\xi)} \right)^{\beta_l} \tag{2.3}$$

(note that a $\Phi_{p,\beta}^b$-optimal design always has a nonsingular information matrix). Thus it is straightforward to verify that the optimal design with respect to the criterion (2.2) proposed by Atkinson and Cox (1974) is the $\Phi_{0,\beta}^b$-optimal discriminating design with respect to the prior $\beta = (\beta_1, \dots, \beta_m)$, where the weights β_l are proportional to $\sum_{j=m-l+1}^{m-1} 1/j, l = 2, \dots, m$ and $\beta_1 = 0$. Note that Atkinson and Cox (1974) considered the problem of discriminating between arbitrary (given) models and that it is straightforward to generalize the optimality criterion on page 327 of their paper to a weighted p-mean in the sense of (2.1). For further results concerning the determination of optimal designs for model discrimination, we refer the reader to the paper of Box and Hill (1967), Atkinson and Fedorov (1975a, b) and Ponce De Leon and Atkinson (1991).

The above definition gives the common c-optimality criterion (for polynomial regression of degree m) setting $\beta_1 = \cdots = \beta_{m-1} = 0$ and $\beta_m = 1$ [see, e.g., Silvey (1980) or Pukelsheim (1981)]. For the choice $b_m = (0, \ldots, 0, 1)' \in \mathbb{R}^{m+1}$ we thus obtain the D_1-optimality criterion considered by Kiefer and Wolfowitz (1959) and Studden (1982b). The above criteria can easily be extended to arbitrary regression functions defined on general design spaces and are extensively discussed in a paper of Dette (1993). In the following discussion we will need the relation between $\Phi^b_{p,\beta}$-optimal designs with respect to different exponents $p < 1$ in (2.1) [for a proof and more details see Dette (1993)].

THEOREM 2.1. *Let $p_1 \in (-\infty, 1)$ and let ξ denote a $\Phi^b_{p_1,\beta}$-optimal discriminating design with respect to the prior β. Then, for every $p_2 \in (-\infty, 1)$, ξ is a $\Phi^b_{p_2,\beta}$-optimal discriminating design with respect to the prior $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_m)$ if and only if*

$$\tilde{\beta}_l = \beta_l \frac{\left(b_l' M_l^{-1}(\xi) b_l\right)^{p_2 - p_1}}{\sum_{j=1}^m \beta_j \left(b_j' M_j^{-1}(\xi) b_j\right)^{p_2 - p_1}}, \qquad l = 1, \ldots, m.$$

3. Optimal discriminating designs. Throughout this paper we assume that the vectors c_l in the definition of $\Phi^c_{p,\beta}$-optimality are given by $b_l = (0, \ldots, 0, 1)' \in \mathbb{R}^{l+1}$, and we will use the criterion (2.1) determining optimal designs for model discrimination. For this task a short description of the theory of canonical moments will be needed and is given in the following. For details the reader is referred to the work of Studden (1980, 1982a, b, 1989), Lau (1983, 1988) and Lim and Studden (1988). Let ξ denote a probability measure on $[a, b]$ with moments $c_i = \int_a^b x^i \, d\xi(x)$. The canonical moments are defined as follows. For a given set of moments $c_0, c_1, \ldots, c_{i-1}$, let c_i^+ denote the maximum of the i-th moment $\int_a^b x^i \, d\mu(x)$ over the set of all probability measures μ having the given moments $c_0, c_1, \ldots, c_{i-1}$. Similarly let c_i^- denote the corresponding minimum. The canonical moments are defined by

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

Note that $0 \leq p_i \leq 1$ and that the canonical moments are left undefined whenever $c_i^+ = c_i^-$. If i is the first index for which this equality holds, then $0 < p_k < 1$, $k = 1, \ldots, i-2$, p_{i-1} must have the value 0 or 1 and the design ξ is supported at a finite number of points [see Skibinsky (1986), Section 1]. The following result will be useful for calculating the support points and weights of a design corresponding to a terminating sequence of canonical moments [see Lau (1988)].

LEMMA 3.1. *Let ξ denote a probability measure on $[a, b]$ with the following canonical moments: $p_{2i} \in (0, 1), i \leq m-1$; $p_{2i-1} = \frac{1}{2}$, $i \leq m$; and $p_{2m} = 1$. Then*

ξ is supported at $m+1$ points $x_0, \dots, x_m$ which are the zeros of the polynomial $(x-b)(x-a)Q_{m-1}(x)$. The masses at the support points are given by

$$\xi(\{x_j\}) = \frac{P_m(x_j)}{\left(d\left[(x-b)(x-a)Q_{m-1}(x)\right]/dx\right)|_{x=x_j}}, \qquad j = 0, \dots, m.$$

Here the polynomials $\{P_l(x)\}_{l=0}^{m}$ and $\{Q_l(x)\}_{l=0}^{m-1}$ are defined recursively by $[P_{-1}(x) = Q_{-1}(x) = 0,\ P_0(x) = Q_0(x) = 1,\ q_j = 1 - p_j,\ j \geq 1]$

$$P_{k+1}(x) = \left(x - \frac{b+a}{2}\right)P_k(x) - \left(\frac{b-a}{2}\right)^2 q_{2k}p_{2k+2}P_{k-1}(x), \qquad k \leq m-1,$$

$$Q_{k+1}(x) = \left(x - \frac{b+a}{2}\right)Q_k(x) - \left(\frac{b-a}{2}\right)^2 p_{2k}q_{2k+2}Q_{k-1}(x), \qquad k \leq m-2.$$

EXAMPLE 3.2. Let $[a, b] = [-1, 1]$ and $m = 3$. Then we have $Q_2(x) = x^2 - p_2q_4$ and $P_3(x) = x^3 - x(q_2p_4 + q_4p_6)$, and it follows by straightforward algebra that the design corresponding to the terminating sequence $(1/2, p_2, 1/2, p_4, 1/2, 1)$ puts masses $p_2p_4/(2(q_2 + p_2p_4))$ at the points ± 1 and masses $q_2/(2(q_2 + p_2p_4))$ at the points $\pm\sqrt{p_2q_4}$.

The following theorem characterizes the $\Phi_{p,\beta}^b$-optimal discriminating design in terms of canonical moments.

THEOREM 3.3. *The $\Phi_{p,\beta}^b$-optimal discriminating design ξ is uniquely determined by its canonical moments $p_1, \dots, p_{2m-1}, p_{2m}$, where*

$$p_{2l-1} = \tfrac{1}{2}, \qquad l = 1, \dots, m, \tag{3.1}$$

$p_{2m} = 1$ and $(p_2, \dots, p_{2m-2})$ is the unique solution of

$$\begin{aligned} &\beta_{l+1}(2p_{2l} - 1)p_{2l+2}^{1+p}\left(\frac{b-a}{2}\right)^{2p} \\ &\quad = \beta_l(1 - p_{2l})^{1-p}(2p_{2l+2} - 1), \qquad l = 1, \dots, m-1. \end{aligned} \tag{3.2}$$

PROOF. Because a $\Phi_{p,\beta}^b$-optimal discriminating design guarantees the estimability of $b_m'a_m$ [i.e., $b_m \in \text{range}(M_m(\xi))$], the optimal design must have a nonsingular information matrix $M_m(\xi)$. From the discussion at the end of Section 2 we have $b_l'M_l^{-1}(\xi)b_l = \det M_{l-1}(\xi)/\det M_l(\xi)$, where the determinant of the information matrix can be expressed in terms of canonical moments, that is ($q_0 = 1$, $q_j = 1 - p_{j,j} \geq 1$),

$$\det M_l(\xi) = (b-a)^{l(l+1)}\prod_{j=1}^{l}(q_{2j-2}p_{2j-1}q_{2j-1}p_{2j})^{l+1-j} \tag{3.3}$$

[see Lau and Studden (1985)]. Observing (3.3) we see that a design maximizing (2.1) must have canonical moments of odd order $p_{2i-1} = \frac{1}{2}$, which corresponds to the symmetry of the design about the point $(a+b)/2$ [see Lau (1983)]. Thus we may restrict ourselves to symmetric designs about the point $(a+b)/2$ and obtain from (3.3), for every symmetric design on $[a, b]$,

$$(3.4) \qquad b_l' M_l^{-1}(\xi) b_l = \left(\frac{2}{b-a}\right)^{2l} \left(\prod_{j=1}^{l} q_{2j-2} p_{2j}\right)^{-1}, \qquad l = 1, \dots, m.$$

In the case $p = 0$ the optimality criterion reduces to the weighted product in (2.3), and it is straightforward to show that the $\Phi_{0,\beta}^{b}$-optimal discriminating design with respect to the prior β is unique and has canonical moments of even order satisfying (3.2) [note that the canonical moments of odd order satisfy (3.1) because of the symmetry of the optimal design]. Conversely, it follows by arguments similar to those given in Dette (1991) that a given design with canonical moments $(\frac{1}{2}, p_2, \frac{1}{2}, \dots, \frac{1}{2}, p_{2m-2}, \frac{1}{2}, 1)$ is $\Phi_{0,\beta}^{b}$-optimal with respect to a prior $\beta^* = (\beta_1^*, \dots, \beta_m^*)$ if and only if

$$\beta_l^* = \left(1 - \frac{q_{2l}}{p_{2l}}\right) \prod_{j=1}^{l-1} \frac{q_{2j}}{p_{2j}}, \qquad l = 1, \dots, m.$$

[Note that this equation defines a one-to-one map from the set of symmetric probability measures supported at a, b and $m-1$ interior points in (a, b) onto the set of all priors β such that there exists a $\Phi_{0,\beta}^{b}$-optimal discriminating design with a nonsingular information matrix $M_m(\xi)$; see Dette (1991).] By Theorem 2.1 [and (3.4)] we obtain that the given design ξ is $\Phi_{p,\beta}^{b}$-optimal with respect to the prior $\beta = (\beta_1, \dots, \beta_m)$ if and only if

$$\begin{aligned} \beta_l &= \beta_l^* \frac{\left(b_l' M_l^{-1}(\xi) b_l\right)^p}{\sum_{l=1}^{m} \beta_l^* \left(b_l' M_l^{-1}(\xi) b_l\right)^p} \\ &= \frac{\left(2/(b-a)\right)^{2lp} \left(\prod_{j=1}^{l} q_{2j-2} p_{2j}\right)^{-p} \prod_{j=1}^{l-1} \left(q_{2j}/p_{2j}\right) \left(1 - q_{2l}/p_{2l}\right)}{\sum_{l=1}^{m} \left(2/(b-a)\right)^{2lp} \left(\prod_{j=1}^{l} q_{2j-2} p_{2j}\right)^{-p} \prod_{j=1}^{l-1} \left(q_{2j}/p_{2j}\right) \left(1 - q_{2l}/p_{2l}\right)}, \end{aligned}$$

$l = 1, \dots, m$. Solving these equations with respect to the canonical moments we have

$$\begin{aligned} \left(\frac{b-a}{2}\right)^{2p} \frac{\beta_{l+1}}{\beta_l} &= \frac{(q_{2l} p_{2l+2})^{-p} \left(q_{2l}/p_{2l}\right)\left(1 - q_{2l+2}/p_{2l+2}\right)}{1 - q_{2l}/p_{2l}} \\ &= \frac{q_{2l}^{1-p}(2p_{2l+2} - 1)}{p_{2l+2}^{1+p}(2p_{2l} - 1)}, \end{aligned}$$

$l = 1, \dots, m-1$, whenever $\beta_l > 0$ and $p_{2l} = \frac{1}{2}$ if $\beta_l = 0$. Thus for a given prior $\beta = (\beta_1, \dots, \beta_m)$ the canonical moments of even order of the $\Phi_{p,\beta}^{b}$-optimal discriminating design with respect to β have to satisfy the equations in (3.2), while the canonical moments of odd order are $\frac{1}{2}$ [i.e., (3.1)] (note that all considered maps are one-to-one). The assertion of the theorem now follows, observing that the equation $\alpha(2x-1) = (1-x)^{1-p}$ has a unique solution in the interval $(0,1)$ whenever $\alpha > 0$ and $p < 1$. □

Theorem 3.3 gives a complete solution of the $\Phi^b_{p,\beta}$-optimal discriminating design problem characterizing the canonical moments of the optimal measure. Note that for the prior $\beta = (0,\dots,0,1)$ the $\Phi^b_{p,\beta}$-optimal discriminating design has canonical moments all equal $\frac{1}{2}$ except p_{2m}, which is 1. In general the support points and the weights of the $\Phi^b_{p,\beta}$-optimal discriminating design have to be calculated by Lemma 3.1. In most cases this has to be done numerically by solving (3.2) recursively [note that $p_{2m} = 1$ and that (3.2) is essentially an equation for only one unknown].

In the following discussion we consider priors $\beta = (\beta_1,\dots,\beta_m)$ such that the ratios β_l/β_{l+1} do not depend on m (e.g., a uniform prior). In this case we can show that there exists a (probability) measure η on $[a,b]$ such that for every $m \in \mathbb{N}$ the $\Phi^b_{p,\beta}$-optimal discriminating design (for polynomial regression up to degree m) is supported at a,b and at the zeros of the $(m-1)$-th orthogonal polynomial with respect to a measure $d\eta(x)$. This can be seen as follows. By Theorem 3.3 the canonical moments of the $\Phi^b_{p,\beta}$-optimal discriminating design ξ^* with respect to the prior β are given by

$$\left(\tfrac{1}{2}, p_2, \tfrac{1}{2}, \dots, \tfrac{1}{2}, p_{2m-2}, \tfrac{1}{2}, 1\right),$$

where the canonical moments of even order p_{2i} are determined by (3.2). Using the theory of continued fractions, it can be shown that the design corresponding to the "reversed" sequence

$$\left(\tfrac{1}{2}, \tilde{p}_2, \tfrac{1}{2}, \dots, \tfrac{1}{2}, \tilde{p}_{2m-2}, \tfrac{1}{2}, 1\right) := \left(\tfrac{1}{2}, 1-p_{2m-2}, \tfrac{1}{2}, \dots, \tfrac{1}{2}, 1-p_2, \tfrac{1}{2}, 1\right)$$

has the same support points as ξ^*[see, e.g., Studden (1982b), Theorem 2.2; or Dette and Studden (1992), Lemma 2.1]. By Lemma 3.1 and Theorem 3.3 these points are given by the zeros of the polynomial $(x-b)(x-a)\tilde{Q}_{m-1}(x)$, where $\tilde{Q}_k(x)$ satisfies the recursive relation $[Q_0(x) = 1,\ \tilde{Q}_{-1}(x) = 0]$

$$\tilde{Q}_{k+1}(x) = \left(x - \frac{b+a}{2}\right)\tilde{Q}_k(x) - \left(\frac{b-a}{2}\right)^2 \tilde{p}_{2k}\tilde{q}_{2k+2}\tilde{Q}_{k-1}(x), \quad k \le m-2, \tag{3.5}$$

and the $\tilde{p}_{2j}$ are defined by $(\tilde{p}_0 = 0)$

$$\beta_{m-j+1}\left(\frac{b-a}{2}\right)^{2p}(2\tilde{p}_{2j}-1)(1-\tilde{p}_{2j-2})^{1+p} = \beta_{m-j}\tilde{p}_{2j}^{1-p}(2\tilde{p}_{2j-2}-1), \tag{3.6}$$

whenever $j \ge 1$. Provided that the ratios β_l/β_{l+1} do not depend on m [as in the case of the uniform prior $\beta^* = (1/m,\dots,1/m)$], the coefficients in the recursion (3.5) defined by (3.6) do not depend on the degree of the polynomial regression m.

Therefore the polynomials $\{\widetilde{Q}_l(x)\}_{l=0}^{\infty}$ defined by (3.5) and (3.6) are orthogonal with respect to a unique probability measure η on the interval $[a, b]$ [see, e.g., Chihara (1978), page 22]. By the results of Studden (1982b) these polynomials are orthogonal with respect to the measure $(b - x)(x - a)\,d\eta(x)$, and we have proved the following theorem, which generalizes Hoel's (1958) famous result for the support of the D-optimal design.

THEOREM 3.4. *Let the ratios of the weights β_l/β_{l+1} be independent of m, $l = 1, \dots, m-1$; let η be the unique measure corresponding to the sequence of canonical moments defined by* (3.6); *and let $\widetilde{Q}_l(x)$ be the lth orthogonal polynomial with respect to the measure $(b - x)(x - a)\,d\eta(x)$. The $\Phi^b_{p,\beta}$-optimal discriminating design with respect to the prior β is supported at the roots of the polynomial $(x - b)(x - a)\widetilde{Q}_{m-1}(x)$.*

EXAMPLE 3.5. Let $[a, b] = [-1, 1]$, let $\beta^* = (1/m, \dots, 1/m)$ denote the uniform prior and let $p = 0$. Then the $\Phi^b_{0,\beta}$-optimality criterion (2.1) reduces to the D-optimality criterion. From (3.6) we obtain $\widetilde{p}_{2j} = j/(2j+1), \widetilde{p}_{2j-1} = \frac{1}{2}$, whenever $j \geq 1$ and η is the uniform distribution on the interval $[-1, 1]$ [see Skibinsky (1969), page 1759]. By Theorem 3.4 the Φ^b_{0,β^*}-optimal discriminating (or D-optimal design) is supported at the zeros of the polynomial $(1 - x^2)\widetilde{Q}_{m-1}(x)$, where $\widetilde{Q}_{m-1}$ is the $(m-1)$th orthogonal polynomial with respect to the measure $(1 - x^2)\,dx$. Observing that $\widetilde{Q}_{m-1}(x)$ must be proportional to the derivative of the mth Legendre polynomial [see, e.g., Abramowitz and Stegun (1964), page 779], we obtain another proof that the D-optimal design is supported at ± 1 and the $m - 1$ zeros of the derivative of the mth Legendre polynomial.

We will conclude this section with the following corollary, which considers the important case of discriminating between polynomial regression models of degree $m - 2, m - 1$ and m. In the following let $T_m(x)$ and $U_m(x)$ denote the mth Chebyshev polynomials of the first and second kind on the interval $[-1, 1]$ orthogonal with respect to the measures $(1 - x^2)^{-1/2}\,dx$ and $(1 - x^2)^{1/2}\,dx$, respectively [see Chihara (1978), pages 1–5].

COROLLARY 3.6. *Let $\beta_m \in (0, 1)$ and $\widehat{\beta} = (0, \dots, 0, 1 - \beta_m, \beta_m)$. Then the $\Phi^b_{0,\widehat{\beta}}$-optimal discriminating design is supported at the points $y_j = ((b - a)x_j + a + b)/2$, $j = 0, \dots, m$, where $x_0, \dots, x_m$ are the zeros of the polynomial*

$$(1 - x^2)\left[U_{m-1}(x) + \alpha U_{m-3}(x)\right]$$

and $0 \leq \alpha < 1$ is the (unique) root of

$$\left(\frac{1 - \alpha}{2}\right)^{1-p} - \frac{\beta_m}{1 - \beta_m}\left(\frac{b - a}{2}\right)^{2p}\alpha = 0. \tag{3.7}$$

The weights at the support points are given by ($x_0 = -1$, $x_m = 1$)

$$\xi(\{y_j\}) = \frac{1-\alpha^2}{(m-1)(1-\alpha^2)+(1+\alpha)^2-4\alpha T^2_{m-1}(x_j)}, \qquad j = 1,\dots,m-1,$$

$$\xi(\{a\}) = \xi(\{b\}) = \frac{\frac{1}{2}(1-\alpha^2)}{(m-1)(1-\alpha^2)+(1-\alpha)^2}.$$

PROOF. From Theorem 3.3 we obtain the following for the canonical moments of the $\Phi^b_{0,\widehat{\beta}}$-optimal discriminating design with respect to the prior $\widehat{\beta}$; $p_j = 1/2, j \le 2m-3, p_{2m-2} = (1+\alpha)/2, p_{2m-1} = 1/2$ and $p_{2m} = 1$, where α is the unique solution of (3.7) in the interval $[0,1)$. Thus the assertion follows directly from the results of Studden (1989) transformed to the interval $[a,b]$. □

REMARK 3.7. It is remarkable that for $[a,b] = [-1,1]$ the $\Phi^b_{0,\beta}$-optimal discriminating design of Corollary 3.6 coincides with the ϕ_p-optimal design (with respect to Kiefer's ϕ_p-criterion) for the highest two coefficients which was determined by Gaffke (1987) (using general results for admissible and invariant designs) and by Studden (1989) (using canonical moments). Note that, contrary to the statement of Corollary 3.6, the results of these authors cannot be transferred to nonsymmetric intervals because the symmetry of the design space is an essential ingredient in their derivation of the ϕ_p-optimal design for the highest two coefficients.

4. Examples and Atkinson and Cox designs. In this section we will present some examples to illustrate the theory given in Section 3.

EXAMPLE 4.1 ($\Phi^b_{p,\beta}$-optimal discriminating designs for cubic regression on the interval $[-1,1]$ with respect to the uniform prior). Let $[a,b] = [-1,1]$ and $\beta^* = (\frac{1}{3},\frac{1}{3},\frac{1}{3})$. We have calculated the Φ^b_{p,β^*}-optimal discriminating designs for cubic regression with respect to the prior β^* by an application of Theorem 3.3 and Lemma 3.1. The results are illustrated in Figures 1 and 2, which show the two positive support points and the corresponding weights of the Φ^b_{p,β^*}-optimal discriminating design with respect to the prior β^* for varying values of p.

In the interval $(-15,-2.5)$ there are no remarkable differences between the $\Phi^b_{p,\beta}$-optimal discriminating designs with respect to different values of $p < -2.5$. Note that we have, for the interval $[-1,1]$,

$$\lim_{p\to-\infty} \Phi^b_{p,\beta^*}(\xi) = \min_{l=1}^{m}\left(b_l' M_l^{-1}(\xi) b_l\right)^{-1} = \frac{\det M_m(\xi)}{\det M_{m-1}(\xi)} \tag{4.1}$$

(independent of the prior β), and thus the $\Phi^b_{p,\beta}$-optimal discriminating design tends to the D_1-optimal designs as $p \to -\infty$ [see Studden (1982b) for more details concerning the D_1-optimal designs in polynomial regression]. For the cubic regression model this design puts masses $\frac{1}{6}, \frac{1}{3}, \frac{1}{3}$ and $\frac{1}{6}$ at the points $-1, -0.5, 0.5$ and 1.

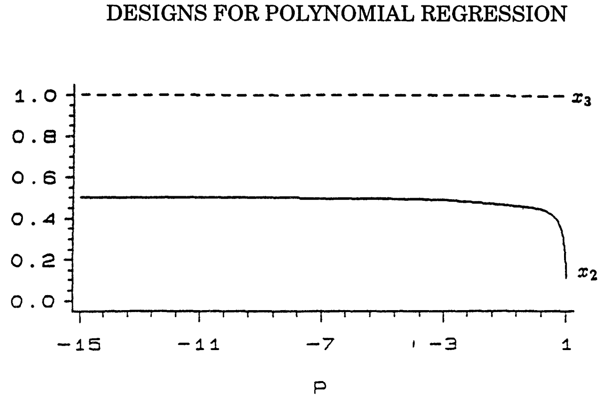

FIG. 1. *Positive support points x_2 and x_3 of the Φ^b_{p,β^*}-optimal discriminating design for $p \in (-15, -1)$.*

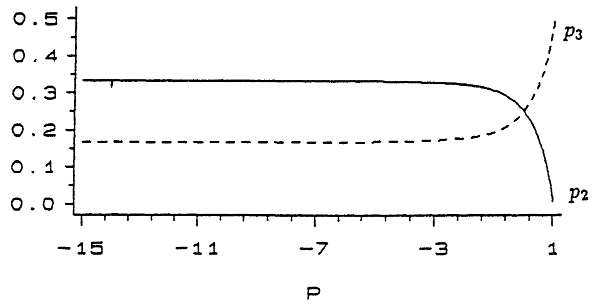

FIG. 2. *Weights p_2 and p_3 corresponding to the positive support points x_2 and x_3 of the Φ^b_{p,β^*}-optimal discriminating design for $p \in (-15, -1)$.*

In the remaining region $[-2.5, 1)$ there are more significant dependencies of the Φ^b_{p,β^*}-optimal discriminating designs from the parameter p. Note here that the case $p = 0$ gives the D-optimal design with equal masses at the points -1, $-1/\sqrt{5}$, $1/\sqrt{5}$ and 1. As p tends to 1 the Φ^b_{p,β^*}-optimal design converges to the design concentrating equal masses at the points -1 and 1. This corresponds to the fact that in the limiting case $p = 1$ a $\Phi^b_{p,\beta}$-optimal discriminating design with respect to any prior $\beta \neq (1, 0, \dots, 0)$ does not exist [for $p = 1$ the solution of (3.2) yields $p_2 = \cdots = p_{2m} = 1$]. Note that there are some similarities concerning the dependency of the optimal designs from p with the results presented by Preitschopf and Pukelsheim (1987), who determined the optimal design for quadratic regression with respect to Kiefer's ϕ_p-criteria.

It might also be of interest how well these discriminating designs do in terms of estimating the parameters in the selected model. To this end we have calculated the D-efficiencies

$$\mathrm{Eff}_l(\xi) := \left[\frac{\det(M_l(\xi))}{\sup_\eta \det(M_l(\eta))}\right]^{1/(l+1)}, \qquad l = 1, 2, 3,$$

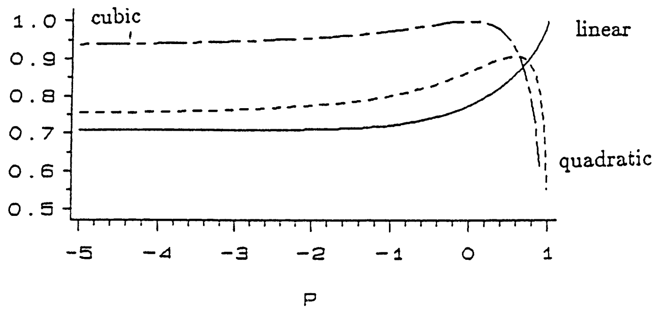

FIG. 3. *D-efficiencies of the Φ^b_{p,β^*}-optimal discriminating design for estimating all parameters in the linear, quadratic and cubic polynomial, $p \in (-5, 1)$.*

for the Φ^b_{p,β^*}-optimal discrimination design for polynomial regression for cubic regression where p varies in the interval $(-5, 1)$. The results are illustrated in Figure 3.

The D-efficiency of the Φ^b_{p,β^*}-optimal discriminating design in the linear model varies between 0.707 ($p = -\infty$) and 1 ($p = 1$), for the quadratic model between 0.0 ($p = 1$) and 0.906 ($p = 0.55$) while for the cubic model we obtain values between 0 ($p = 1$) and 1 ($p = 0$). From Figure 3 we see that the Φ^b_{p,β^*}-optimal discriminating design has good efficiencies for the estimation of the parameters in the selected model if $p \in [-1, 0.5]$. For example, if $p = 0.5$, we obtain for the D-efficiency in the linear, quadratic and cubic model 0.8414, 0.9049 and 0.9548, respectively. Of course, these results will vary with different priors β and different regression intervals $[a, b]$.

EXAMPLE 4.2 (Change of the prior β and the interval $[a, b]$). Obviously, different priors and different intervals $[a, b]$ yield different canonical moments of the $\Phi^b_{p,\beta}$-optimal discriminating design (except in the case $p = 0$), but it is remarkable that the behavior of the $\Phi^b_{p,\beta}$-optimal discriminating designs at the limiting points $p = -\infty$ and $p = 1$ described in Example 4.1 depends essentially on the prior β and on the interval $[a, b]$. To see this we consider at first the cubic regression on the interval $[-1, 1]$ and the prior $\gamma = \frac{1}{14}(2, 4, 8)$. Observing (3.2) (for the case $p = 1$), it is straightforward to show that the $\Phi^b_{1,\gamma}$-optimal discriminating design with respect to the prior γ has canonical moments $(\frac{1}{2}, \frac{13}{18}, \frac{1}{2}, \frac{3}{4}, \frac{1}{2}, 1)$. From Example 3.2 we have that the optimal measure puts masses $\frac{39}{118}, \frac{10}{59}, \frac{10}{59}$ and $\frac{39}{118}$ at the points $-1, -\sqrt{13/72}, \sqrt{13/72}$ and 1 while, for the uniform prior β^*, a Φ^b_{1,β^*}-optimal discriminating design on the interval $[-1, 1]$ does not exist. More generally, it is easy to see that the $\Phi^b_{1,\gamma}$-optimal discriminating design with respect to a prior β satisfying $\beta_l < \beta_{l+1}$, $l = 1, \dots, m-1$, exists for polynomial regression of arbitrary degree m on the interval $[-1, 1]$ while the Φ^b_{1,β^*}-optimal discriminating design with respect to the uniform prior fails to exist on this interval.

TABLE 1

Support points and weights of the optimal discriminating designs of Atkinson and Cox for polynomial regression models up to degree 8: Left column, maximum degree of the polynomials; first line, support points; second line, weights

1	−1.0000	1.0000							
	0.5000	0.5000							
2	−1.0000	0.0000	1.0000						
	0.2500	0.5000	0.2500						
3	−1.0000	−0.4629	0.4629	1.0000					
	0.1818	0.3181	0.3181	0.1818					
4	−1.0000	−0.6715	0.0000	0.6715	1.0000				
	0.1428	0.2376	0.2392	0.2376	0.1428				
5	−1.0000	−0.7795	−0.2926	0.2926	0.7795	1.0000			
	0.1176	0.1908	0.1916	0.1916	0.1908	0.1176			
6	−1.0000	− 0.8422	−0.4785	0.0000	0.4785	0.8422	1.0000		
	0.1000	0.1598	0.1601	0.1602	0.1598	0.1600	0.1000		
7	−1.0000	− 0.8815	−0.6016	−0.2133	0.2133	0.6016	0.8815	1.0000	
	0.0870	0.1376	0.1377	0.1377	0.1377	0.1377	0.1376	0.0870	
8	−1.0000	−0.9079	−0.6864	−0.3689	0.0000	0.3689	0.6864	0.9079	1.0000
	0.0769	0.1209	0.1209	0.1209	0.1208	0.1209	0.1209	0.1209	0.0769

Similarly, the limiting behavior of the $\Phi^b_{p,\beta}$-optimal discriminating design as p converges to $-\infty$ is changing with the underlying interval $[a, b]$ of the polynomial regression. In this case the relation (4.1) is not necessarily true any longer [this follows readily from (3.4)]. As an example we consider the Φ^b_{p,β^*}-optimal discriminating design with respect to the uniform prior for the cubic regression on the interval $[-5, 5]$. By numerical computations we obtain that the Φ^b_{p,β^*}-optimal discriminating design converges to the design with masses 0.4778, 0.0222, 0.0222 and 0.4778 at the points $-5, -0.9889, 0.9889$ and 5 as p tends to $-\infty$. This design is not the D_1-optimal design for cubic regression on the interval $[-5, 5]$ which puts masses $\frac{1}{6}, \frac{1}{3}, \frac{1}{3}$ and $\frac{1}{6}$ at the points $-5, -2.5, 2.5$ and 5.

EXAMPLE 4.3 (The discriminating designs of Atkinson and Cox). In this example we will determine the optimal discriminating designs for polynomial regression models of degree $l = 1, \dots, m$ in the sense of Atkinson and Cox (1974). By the discussion of Section 2 we have to find the $\Phi^b_{0,\beta}$-optimal discriminating designs with respect to the prior $\widehat{\beta} = (\widehat{\beta}_1, \dots, \widehat{\beta}_m)$, where $\widehat{\beta}_l$ is proportional to k[4]$\Sigma_{j=m-l+1}^{m-1}(1/j)$, $l = 2, \dots, m$ and $\widehat{\beta}_1 = 0$. We have calculated the optimal designs for $m = 2, \dots, 8$ on the interval $[-1, 1]$ (note that in the case $p = 0$ the $\Phi^b_{0,\beta}$-optimal designs on other intervals $[a, b]$ can easily be obtained by a linear transformation of the optimal design on $[-1, 1]$). The results are given in Table 1 and indicate that the optimal design (i.e., the $\Phi^b_{0,\beta}$-optimal discriminating design with respect to the prior $\widehat{\beta}$) converges to a design with equal masses at the interior support points when the degree of the regression m tends to infinity.

THEOREM 4.4. *Let* $\widehat{\beta} = (0, \widehat{\beta}_2, \dots, \widehat{\beta}_m)$, *where* $\widehat{\beta}_l$ *is proportional to* $\sum_{j=m-l+1}^{m-1} (1/j)$, $l = 2, \dots, m$. *Then the* $\Phi^b_{0,\widehat{\beta}}$-*optimal discriminating design with respect to the prior* $\widehat{\beta}$ *on the interval* $[a,b]$ *converges weakly to the arcsine measure with density proportional to* $(b-x)^{-1/2}(x-a)^{-1/2}\,dx$.

PROOF. Applying Theorem 3.3 (for $p = 0$), it is straightforward to show that the $\Phi^b_{0,\hat{\beta}}$-optimal discriminating design with respect to the prior $\widehat{\beta}$ has canonical moments $p_{2l-1} = \frac{1}{2}$, $l = 1, \dots, m$, $p_{2m} = 1$, $p_2 = \frac{1}{2}$ and $p_{2l} = \sigma_l/(\sigma_l + \sigma_{l+1})$, $l = 2, \dots, m-1$, where the numbers σ_l are given by

$$\sigma_l = (m - l + 1)\left\{1 + \frac{1}{m-1} + \cdots + \frac{1}{m-l+2}\right\}, \qquad l = 2, \dots, m-1.$$

Thus we have $\lim_{m \to \infty} p_l = \frac{1}{2}$, and the assertion follows because the arcsine distribution is the only distribution having $p_l = \frac{1}{2}$ for all $l \in \mathbb{N}$ [see Skibinsky (1969)]. □

The proof of Theorem 4.4 shows that the canonical moments of the optimal design converge to $\frac{1}{2}$ when $m \to \infty$. Thus, for large m, the discriminating design of Atkinson and Cox (1974) for polynomial regression models of degree $l = 1, \dots, m$ can be approximated by the design with masses $1/(2m)$ at the points -1 and 1 and masses $1/m$ at the zeros of the polynomial $U_{m-1}((2x-a-b)/(b-a))$ which has the canonical moments $p_l = \frac{1}{2}$, $l \le 2m-1$ and $p_{2m} = 1$ and is in fact the D_1-optimal design for polynomial regression of degree m [see, e.g., Studden (1982b)].

Acknowledgments. This paper was inspired by a discussion during a workshop on optimal design in Irsee (1992). The author would like to thank Professor K. Chaloner for proposing the problem of model discrimination and Professor N. Gaffke and Professor F. Pukelsheim for organizing this nice and interesting conference.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions*. Dover, New York.

ATKINSON, A. C. (1972). Planning experiments to detect inadequate regression models. *Biometrika* **59** 275–293.

ATKINSON, A. C. and COX, D. R. (1974). Planning experiments for discriminating between models (with discussion). *J. Roy. Statist. Soc. Ser. B* **36** 321–348.

ATKINSON, A. C. and FEDOROV, V. V. (1975a). The design of experiments for discriminating between two rival models. *Biometrika* **62** 57–70.

ATKINSON, A. C. and FEDOROV, V. V. (1975b). Optimal design: experiments for discriminating between several models. *Biometrika* **62** 289–303.

BOX, G. E. P. and HILL, W. J. (1967). Discrimination among mechanistic models. *Technometrics* **9** 57–71.

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

DETTE, H. (1991). A note on model robust design for polynomial regression. *J. Statist. Plann. Inference* **28** 223–232.

DETTE, H. (1993). A new interpretation of optimality for E-optimal designs in linear regression models. *Metrika* **40** 37–50.

DETTE, H. and STUDDEN, W. J. (1992). A new characterization of the classical orthogonal polynomials. *J. Approx. Theory* **53** 1–17.

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

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

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

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

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

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

LAU, T. S. (1988). D-optimal designs on the unit q-ball. *J. Statist. Plann. Inference* **19** 299–315.

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

LIM, Y. B. and STUDDEN, W. J. (1988). Efficient D_s-optimal designs for multivariate polynomial regression on the q-cube. *Ann. Statist.* **16** 1225–1240.

MURTY, V. N. and STUDDEN, W. J. (1972). Optimal designs for estimating the slope of a polynomial regression. *J. Amer. Statist. Assoc.* **67** 869–873.

PONCE DE LEON, A. C. and ATKINSON, A.C. (1991). Optimal experimental design for discrimination between two rival models in the presence of prior information. *Biometrika* **78** 601–608.

PREITSCHOPF, F. and PUKELSHEIM, F. (1987). Optimal designs for quadratic regression. *J. Statist. Plann. Inference* **16** 213–218.

PUKELSHEIM, F. (1981). On c-optimal design measures. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 13–20.

SILVEY, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

SKIBINSKY, M. (1969). Some striking properties of binomial and beta moments. *Ann. Math. Statist.* **40** 1753–1764.

SKIBINSKY, M. (1986). Principal representations and canonical moment sequences for distributions on an interval. *J. Math. Anal. Appl.* **120** 95–120.

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

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

STUDDEN, W. J. (1982b). 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.

STUDDEN, W. J. (1989). Note on some Φ_p-optimal designs for polynomial regression. *Ann. Statist.* **17** 618–623.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
TECHNISCHE UNIVERSITÄT DRESDEN
MOMMSENSTR. 13
01062 DRESDEN
GERMANY