OPTIMAL DESIGNS FOR IDENTIFYING THE DEGREE OF A POLYNOMIAL REGRESSION[1]

BY HOLGER DETTE

Technische Universität Dresden

If an experimenter wants to determine the degree of a polynomial regression on the basis of a sample of observations, Anderson showed that the following method is optimal. Starting with the highest (specified) degree the procedure is to test in sequence whether the coefficients are 0. In this paper optimal designs for Anderson's procedure are determined explicitly. The optimal design maximizes the minimum power of a given set of alternatives.

1. Introduction. A frequent problem in regression analysis is to determine how many independent variables have to be included in the fitted regression function. In many cases (e.g., polynomial regression) the underlying models are nested

$$(1.1) \qquad h_l(x) = \sum_{i=0}^{l} \vartheta_i f_i(x), \qquad l = 0, \ldots, n,$$

and based on a sample of observations the experimenter has to identify the appropriate model $h_l(x)$. Anderson (1962) studied the following decision rule. For a given set of levels $(\alpha_1, \ldots, \alpha_n)$ the procedure chooses the largest integer in $\{1, \ldots, n\}$, for which the F-test in the model $h_j(x)$ rejects the hypotheses H_0: $\vartheta_j = 0$ at the levels α_j, $j = 1, \ldots, n$. It is well known that Anderson's method has several optimality properties [see Anderson (1962) or Spruill (1990)]. In the following, $Y_1, \ldots, Y_m$ denote m independent normally distributed observations with common variance $\sigma^2 > 0$ and mean given by one of the models in (1.1), that is,

$$Y = Z_l \theta_l + \varepsilon \quad \text{for some } l = 0, \ldots, n,$$

where $Y = (Y_1, \ldots, Y_m)'$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_m)$, $\theta_l = (\vartheta_0, \ldots, \vartheta_l)'$ and $Z_l = (f_j(x_i))_{i=0,\ldots,m}^{j=0,\ldots,l}$, $l = 0, \ldots, n$. For Anderson's procedure the probability of the error of choosing too many functions is independent of the matrix Z_l provided that Z_l has full rank [see Anderson (1962)]. If the model is $h_l(x) = h_{l-1}(x) +$

1248

$\vartheta_l f_l(x)$, the distribution of the test statistic of the F-test for the hypothesis H_0: $\vartheta_l = 0$ has the noncentrality parameter

$$(1.2) \qquad \delta_l^2 = \frac{\vartheta_l^2}{\sigma^2} \left(e_l'(Z_l'Z_l)^{-1} e_l \right)^{-1},$$

where $e_l = (0, \ldots, 0, 1)' \in \mathbb{R}^{l+1}$ [see Pukelsheim (1993), page 70]. Consequently, the probability of deciding in favor of $h_{l-1}(x)$ [when the model is in fact $h_l(x)$] is a decreasing function of δ_l^2, and a good choice of a design $x_0, \ldots, x_m$ will make the quantities in (1.2) as large as possible.

In a recent paper Spruill (1990) considered similar problems and determined the optimal approximate design with respect to a maximin criterion which maximizes the local power of the F-tests. In this paper we investigate a different maximin criterion which depends directly on the noncentrality parameters in testing the degree of a polynomial regression using Anderson's method. Therefore the structure of the optimal design changes completely when the observations can be taken in a different design space or different alternatives are assumed. In Section 2 we describe some general aspects of (approximate) design theory and introduce the optimality criterion. Section 3 gives a short review of the theory of canonical moments which were introduced by Studden (1980, 1982a, b) in the context of design theory. These results are applied in Sections 3 and 4 in order to obtain a complete solution of the proposed optimal design problem. In the special case $[a, b] = [-1, 1]$, $\vartheta_l = 1$, $l = 1, \ldots, n$, the design with respect to the proposed criterion coincides with the locally maximin design given in Spruill (1990). Finally, some asymptotic considerations and examples are presented in Section 5.

2. The optimality criterion.

Let $\mathscr{X}$ denote a compact space with a sigma field containing all one-point sets and at least $n + 1$ points. In the following we consider $n + 1$ linearly independent, real-valued and continuous regression functions $f_0(x), \ldots, f_n(x)$ (defined on $\mathscr{X}$) and collect the first $l + 1$ functions in a vector $g_l(x) = (f_0(x), \ldots, f_l(x))'$, $l = 0, \ldots, n$. The model in (1.1) can now be written as $h_l(x) = \theta_l' g_l(x)$. An (approximate) design ξ is a probability measure on $\mathscr{X}$ and the matrix

$$M_l(\xi) = \int_{\mathscr{X}} g_l(x) g_l'(x) \, d\xi(x)$$

is called the information matrix for the model $h_l(x)$, $l = 0, \ldots, n$. If ξ puts masses m_i/m, $i = 1, \ldots, s$, at the points $x_1, \ldots, x_s$, the experimenter takes m uncorrelated observations, m_i at each x_i, $i = 1, \ldots, s$, and the inverse of the information matrix $M_l(\xi)$ (in the model h_l) is proportional to the covariance matrix of the least squares estimator for θ_l, that is, $\sigma^2(Z_l'Z_l)^{-1}$. According to the discussion following (1.2), we call an approximate design ξ (maximin) *optimal discriminating design* if ξ maximizes the function

$$(2.1) \qquad \Phi(\xi) = \min\left\{ \vartheta_l^2 \left(e_l' M_l^{-1}(\xi) e_l \right)^{-1} \middle| l = 1, \ldots, n \right\}.$$

Note that this optimality criterion is a local criterion in the sense that it depends on the alternatives ϑ_l of the corresponding hypotheses H_0: $\vartheta_l = 0$, $l = 1, \ldots, n$. Spruill's (1990) locally maximin designs on the interval $[-1, 1]$ agree with the optimal discriminating designs for the alternatives $\vartheta_1 = \cdots = \vartheta_n = 1$, and we will discuss a generalization of this special choice in the following sections. An extremely useful tool for determining optimal (approximate) designs is the concept of equivalence theorems which provides necessary and sufficient conditions for a design to be optimal [see, e.g., Pukelsheim (1993)]. In order to derive such a result for the optimality criterion (2.1), we define, for a design ξ with nonsingular information matrix $M_n(\xi)$,

$$\tilde{M}(\xi) = \begin{pmatrix} M_1(\xi) & & \\ & \ddots & \\ & & M_n(\xi) \end{pmatrix} \in \mathbb{R}^{n(n+3)/2 \times n(n+3)/2},$$

$$K = \begin{pmatrix} \vartheta_1^{-1}e_1 & & \\ & \ddots & \\ & & \vartheta_n^{-1}e_n \end{pmatrix} \in \mathbb{R}^{n(n+3)/2 \times n}$$

and

$$C_K(\tilde{M}) = \left(K'\tilde{M}(\xi)^{-1}K \right)^{-1}$$

$$= \begin{pmatrix} \vartheta_1^2\left(e_1'M_1^{-1}(\xi)e_1\right)^{-1} & & \\ & \ddots & \\ & & \vartheta_n^2\left(e_n'M_n^{-1}(\xi)e_n\right)^{-1} \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

[Note that all other entries in these matrices are zero and that an optimal discriminating design must have nonsingular matrices $M_1(\xi), \ldots, M_n(\xi)$.] If an information function j is defined on the nonnegative definite $n \times n$ matrices by $j(A) = \lambda_{\min}(A)$ [here $\lambda_{\min}(A)$ denotes the minimum eigenvalue of A], then it follows for criterion (2.1) that

$$\Phi(\xi) = j\left(C_K(\tilde{M}(\xi)) \right), \qquad \det M_n(\xi) > 0.$$

By an application of the general equivalence theorem in Pukelsheim [(1993), page 175] it is now straightforward to show the following equivalence theorem for the optimality criterion defined in (2.1).

THEOREM 2.1. *Let ξ denote a design such that $M_n(\xi)$ is nonsingular, and let $\mathscr{N}(\xi) := \{j|\vartheta_j^2(e_j'M_j^{-1}(\xi)e_j)^{-1} = \min_{l=1}^n \vartheta_l^2(e_l'M_l^{-1}(\xi)e_l)^{-1}\}$. The design ξ*

is an optimal discriminating design if and only if there exist nonnegative numbers α_l, $l \in \mathcal{N}(\xi)$, with $\sum_{l \in \mathcal{N}(\xi)} \alpha_l = 1$ such that

$$(2.2) \qquad \sum_{l \in \mathcal{N}(\xi)} \alpha_l \frac{\left(e_l' M_l^{-1}(\xi) g_l(x)\right)^2}{e_l' M_l^{-1}(\xi) e_l} \leq 1,$$

for all $x \in \mathscr{X}$. Moreover, in (2.2) equality holds for all support points of every optimal discriminating design.

REMARK 2.2. Theorem 2.1 can easily be generalized by changing the e_l to arbitrary vectors $c_l \in \mathbb{R}^{l+1}$ and not necessarily nested regression functions. In this case the design ξ has to satisfy $c_l \in \mathrm{range}(M_l(\xi))$, $l = 1, \ldots, n$, and the inverses in Theorem 2.1 have to be replaced by general inverses [see Pukelsheim (1993), page 283].

3. Polynomial regression models. Let $g_l(x) = (1, x, \ldots, x^l)'$ denote the vector of monomials up to the order l and $\mathscr{X} = [a, b]$. Thus the problem of Sections 1 and 2 is to determine the optimal design when Anderson's procedure is applied for testing the degree of a polynomial regression. For this purpose we need some basic facts about canonical moments which were introduced in the context of design theory by Studden (1980, 1982a, b) [see also Lau (1983) for more details]. Let ξ denote a probability measure on $[a, b]$ with moments $c_i = \int_a^b x^i \, d\xi(x)$. For a given set of moments $c_0, \ldots, c_{i-1}$ let c_i^+ denote the maximum of the ith moment $\int_a^b x^i \, d\xi(x)$ over the set of all probability measures μ on $[a, b]$ having the given moments $c_0, c_1, \ldots, c_{i-1}$. Similarly let c_i^- denote the corresponding minimum. The ith canonical moment is 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. In this case ξ is the "lower" or "upper principal representation" of its corresponding moment point $(c_0, \ldots, c_{i-1})$ [see Skibinsky (1986), Section 1, or Chang, Kemperman and Studden (1993)]. The optimality criterion (2.1) can easily be expressed in terms of canonical moments [see, e.g., Lau and Studden (1985)]:

$$(3.1) \qquad \begin{aligned} \Phi(\xi) &= \min\left\{ \vartheta_l^2 \frac{\det M_l(\xi)}{\det M_{l-1}(\xi)} \,\middle|\, l = 1, \ldots, n \right\} \\ &= \min\left\{ \vartheta_l^2 (b - a)^{2l} \prod_{j=1}^{l} q_{2j-2} p_{2j-1} q_{2j-1} p_{2j} \,\middle|\, l = 1, \ldots, n \right\} \end{aligned}$$

(here $q_j = 1 - p_j$, $j \geq 1$, $q_0 = 1$). A probability measure ξ is symmetric about the point $(a + b)/2$ if and only if all canonical moments of odd order satisfy $p_{2i-1} = \frac{1}{2}$.

The following result shows that every symmetric design maximizes a (weighted) geometric mean of the ratios $\det M_l(\xi)/\det M_{l-1}(\xi)$ and can be proved by similar arguments as in Dette (1991).

THEOREM 3.1. *Let ξ^* denote a design with canonical moments $(\frac{1}{2}, p_2^*, \frac{1}{2}, p_4^*, \ldots, \frac{1}{2}, p_{2n-2}^*, \frac{1}{2}, 1)$ and define*

$$(3.2) \qquad \beta_l^* = \left(1 - \frac{q_{2l}^*}{p_{2l}^*}\right) \prod_{j=1}^{l-1} \frac{q_{2j}^*}{p_{2j}^*}, \qquad l = 1, \ldots, n.$$

Then the design ξ^* maximizes the weighted geometric mean

$$\prod_{j=1}^{n} \left(\frac{\det M_j(\xi)}{\det M_{j-1}(\xi)}\right)^{\beta_j^*}$$

among all designs on the interval $[a, b]$.

If ξ has canonical moments $(p_1, p_2, p_3, p_4, p_5, \ldots)$, then its reflection ξ' about the point $(a + b)/2$ has canonical moments $(q_1, p_2, q_3, p_4, q_5, \ldots)$ [see, e.g., Lau and Studden (1985), page 387]. Consequently, we obtain from (3.1) that $\Phi(\xi) = \Phi(\xi')$, and it follows by standard arguments of optimal design theory that the optimal discriminating design is symmetric about the point $(a + b)/2$.

THEOREM 3.2. *The optimal discriminating design ξ^* has canonical moments $\frac{1}{2}, p_2^*, \frac{1}{2}, p_4^*, \frac{1}{2}, \ldots, p_{2n-2}^*, \frac{1}{2}, 1$, where the canonical moments of even order are defined for $j = 1, \ldots, n - 1$ recursively by $(p_{2n}^* = 1)$*

$$(3.3) \quad p_{2(n-j)}^* = \max\left\{1 - \frac{\vartheta_{n-j}^2}{\vartheta_n^2} \left(\frac{b - a}{2}\right)^{-2j} \prod_{i=n-j+1}^{n-1} (q_{2i}^* p_{2i}^*)^{-1}, \frac{1}{2}\right\}.$$

PROOF. Consider the design ξ^* with canonical moments $(\frac{1}{2}, p_2^*, \frac{1}{2}, \ldots, \frac{1}{2}, p_{2n-2}^*, \frac{1}{2}, 1)$ defined by (3.3), and let

$$\gamma_j := 1 - \frac{\vartheta_{n-j}^2}{\vartheta_n^2} \left(\frac{b - a}{2}\right)^{-2j} \left(\prod_{i=n-j+1}^{n-1} q_{2i}^* p_{2i}^*\right)^{-1}, \qquad j = 1, \ldots, n - 1.$$

If $\gamma_j \geq \frac{1}{2}$, then it is easy to see [observing (3.1) and (3.3)] that

$$\vartheta_{n-j}^2 \frac{\det M_{n-j}(\xi^*)}{\det M_{n-j-1}(\xi^*)} = \vartheta_n^2 \frac{\det M_n(\xi^*)}{\det M_{n-1}(\xi^*)}.$$

On the other hand, if $\gamma_j < \frac{1}{2}$, then $p^*_{2(n-j)} = q^*_{2(n-j)} = \frac{1}{2}$ and it follows that $(1 - \gamma_j)^{-1} < 2$, which implies

$$\frac{\vartheta_n^2}{\vartheta_{n-j}^2} \frac{\det M_n(\xi^*)}{\det M_{n-1}(\xi^*)} \frac{\det M_{n-j-1}(\xi^*)}{\det M_{n-j}(\xi^*)} = \frac{q^*_{2(n-j)}}{1 - \gamma_j} < 1.$$

Consequently, the set $\mathscr{N}(\xi^*)$ defined in Theorem 2.1 is given by ($\gamma_0 := 1$)

(3.4) $\mathscr{N}(\xi^*) = \left\{ j \in \{1, \ldots, n\} \, \big| \, \gamma_{n-j} \geq \frac{1}{2} \right\}.$

By Theorem 3.1 the design ξ^* maximizes the weighted geometric mean

$$\prod_{l=1}^{n} \left(\frac{\det M_l(\xi)}{\det M_{l-1}(\xi)} \right)^{\beta_l^*},$$

where the weights β_l^* are defined in (3.2). If all weights satisfy $\beta_l^* \geq 0$, $l = 1, \ldots, n$, then Dette [(1993), Theorem 2.2] shows that this property is equivalent to the condition

(3.5) $\displaystyle\sum_{l=1}^{n} \beta_l^* \frac{\left(e_l' M_l^{-1}(\xi^*) g_l(x) \right)^2}{e_l' M_l^{-1}(\xi^*) e_l} \leq 1,$

for all $x \in [a, b]$. If $l \notin \mathscr{N}(\xi^*)$, then we have $\gamma_{n-l} < \frac{1}{2}$, by (3.3), $p^*_{2l} = \frac{1}{2}$ and by (3.2) this implies $\beta_l^* = 0$. On the other hand, if $\gamma_{n-l} \geq \frac{1}{2}$, it follows from $p^*_{2l} \geq \frac{1}{2}$ that $\beta_l^* \geq 0$. Therefore Theorem 2.2 in Dette (1993) can be applied and (3.5) may be rewritten as

$$\sum_{l \in \mathscr{N}(\xi^*)} \beta_l^* \frac{\left(e_l' M_l^{-1}(\xi^*) g_l(x) \right)^2}{e_l' M_l^{-1}(\xi^*) e_l} \leq 1,$$

for all $x \in [a, b]$. The assertion follows now from Theorem 2.1. $\square$

The weights and the support points of the optimal discriminating design ξ^* corresponding to the terminating sequence of canonical moments can be calculated by standard techniques [see Lau (1988)]. Define $\sigma = (b - a)/2$, $\tau = (a + b)/2$,

$$K\begin{pmatrix} & a_1 & & a_n & \\ b_0 & b_1 & \cdots & & b_n \end{pmatrix} = \det \begin{pmatrix} b_0 & -1 & & & & \\ a_1 & b_1 & -1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & b_{n-1} & -1 \\ & & & & a_n & b_n \end{pmatrix}$$

(all other entries in the matrix are 0) and the polynomials

(3.6) $\displaystyle P_n(x) = K\begin{pmatrix} -\sigma^2 q^*_{2n-2} p^*_{2n} & \cdots & -\sigma^2 q^*_2 p^*_4 \\ x - \tau \quad x - \tau & \cdots & x - \tau \quad x - \tau \end{pmatrix},$

(3.7) $\displaystyle Q_{n-1}(x) = K\begin{pmatrix} -\sigma^2 p^*_{2n-4} q^*_{2n-2} & \cdots & -\sigma^2 p^*_2 q^*_4 \\ x - \tau \quad x - \tau & \cdots & x - \tau \quad x - \tau \end{pmatrix}$

$[P_0(x) = Q_0(x) = 1, \; P_1(x) = Q_1(x) = x - \tau]$, then the following result holds [see, e.g., Lau (1988)].

PROPOSITION 3.3.   *The optimal discriminating design* ξ^* *is supported at the* $n + 1$ *zeros* $x_0, \ldots, x_n$ *of the polynomial* $(x - a)(x - b)Q_{n-1}(x)$ *with masses*

$$\xi^*(\{x_j\}) = \frac{P_n(x_j)}{d/dx[(x - a)(x - b)Q_{n-1}(x)]|_{x = x_j}}, \qquad j = 0, \ldots, n.$$

EXAMPLE 3.4.   Let $a = -b = -1$ and $n = 2$ (quadratic regression), by Theorem 3.2 the optimal discriminating design has canonical moments (of even order) $p_4 = 1$ and $p_2 = \max\{1 - \vartheta_1^2/\vartheta_2^2, 1/2\}$. Therefore, if $|\vartheta_1|$ is small in proportion to $|\vartheta_2|$ (i.e., $|\vartheta_1| \le |\vartheta_2|/\sqrt{2}$), the optimal design ξ_1^* puts masses $1/2 - \vartheta_1^2/(2\vartheta_2^2), \vartheta_1^2/\vartheta_2^2, 1/2 - \vartheta_1^2/(2\vartheta_2^2)$ at the points $-1, 0, 1$. This corresponds to the somewhat intuitive fact that one only needs a few observations at zero in order to distinguish between a linear or quadratic regression on $[-1, 1]$ when it is known that the extremum of the quadratic function is attained in a neighborhood of 0. On the other hand, if $|\vartheta_2|$ is small compared to $|\vartheta_1|$ (i.e., $|\vartheta_1| > |\vartheta_2|/\sqrt{2}$), the optimal discriminating design ξ_2^* has $p_2^* = \frac{1}{2}$ and masses $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ at $-1, 0, 1$. In this case the minimum of the quadratic polynomial is attained outside of the interval $[-1, 1]$ and the linear and quadratic functions have a similar form inside $[-1, 1]$. In order to distinguish between these models, one has to take the best design for testing the highest coefficient of the quadratic regression which is given by ξ_2^* [see also Kiefer and Wolfowitz (1959)].

4. Explicit solutions.   Throughout this and the following section we consider the model of Section 3 with a special set of alternatives. More precisely, we will assume that the unknown parameters in criterion (2.1) satisfy $\vartheta_l = \vartheta^l$, $l = 1, \ldots, n$, for some $\vartheta > 0$. For this case criterion (2.1) reduces to [observing (3.1) and the fact that the optimal discriminating design is symmetric, which means $p_{2j-1} = \frac{1}{2}$, $j = 1, \ldots, n$]

$$\Phi(\xi) = \min\left\{\vartheta^{2l} \left. \frac{\det M_l(\xi)}{\det M_{l-1}(\xi)} \right| l = 1, \ldots, n\right\}$$

(4.1)

$$= \min\left\{\rho^{2l} \prod_{j=1}^{l} q_{2j-2} p_{2j} \middle| l = 1, \ldots, n\right\},$$

where $\rho = \vartheta(b - a)/2$. An optimal design with respect to the criterion (4.1) maximizes the noncentrality parameters of the corresponding F-distributions assuming that the alternatives ϑ_l of the hypotheses $H_l: \vartheta_l = 0$ are of the form $\vartheta_l = \vartheta^l$, $l = 1, \ldots, n$, for some $\vartheta > 0$.

The following result gives the canonical moments of the optimal design with respect to this criterion. Here and throughout this paper $U_n(x)$ will

denote the Chebyshev polynomial of the second kind [see, e.g., Rivlin (1990), page 7].

THEOREM 4.1. *The optimal discriminating design* [*with respect to criterion* (4.1)] *has canonical moments* $(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, p_{2k}^*, \frac{1}{2}, \ldots, \frac{1}{2}, p_{2n-2}^*, \frac{1}{2}, 1)$, *where*

$$(4.2) \qquad p_{2j}^* = \frac{U_{n-j+1}(\rho/2)}{\rho U_{n-j}(\rho/2)}, \qquad j = n, n-1, \ldots, k,$$

and

$$(4.3) \qquad k = \min\left\{ j \leq n \left| \frac{U_{n-i+1}(\rho/2)}{\rho U_{n-i}(\rho/2)} > \frac{1}{2} \text{ for } i = j, \ldots, n \right. \right\}.$$

PROOF. Let $\varepsilon_n = 1$ and consider the sequence

$$(4.4) \qquad \varepsilon_j = 1 - \frac{\rho^{-2}}{\varepsilon_{j+1}}, \qquad j = n-1, \ldots, 1.$$

It is straightforward to show that $\varepsilon_j \leq \varepsilon_{j+1}$ whenever $\varepsilon_n, \ldots, \varepsilon_{j+1} > 0$ and obviously $\varepsilon_n = 1 = U_1(\rho/2)/\rho U_0(\rho/2)$ [see Rivlin (1990), page 39]. In the following we show that ε_j is given by the right-hand side of (4.2) for all $j = 1, \ldots, n$. To this end we use the recursive relation for the Chebyshev polynomials of the second kind,

$$U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x), \qquad U_0(x) = 1, U_{-1}(x) = 0$$

[see Rivlin (1990), page 40] and obtain by induction $(j + 1 \to j)$ that

$$(4.5) \quad \varepsilon_j = 1 - \frac{\rho^{-2}\rho U_{n-j-1}(\rho/2)}{U_{n-j}(\rho/2)} = \frac{U_{n-j+1}(\rho/2)}{\rho U_{n-j}(\rho/2)}, \qquad j = 1, \ldots, n.$$

By Theorem 3.2 the canonical moments of even order of the optimal discriminating design [with respect to criterion (4.1)] are given by

$$(4.6) \qquad p_{2j}^* = \max\left\{ 1 - \rho^{-2(n-j)} \prod_{i=j+1}^{n-1} (q_{2i}^* p_{2i}^*)^{-1}, \frac{1}{2} \right\}.$$

If $\rho \leq 1$, then we have $p_{2j}^* = \frac{1}{2}$, $j = 1, \ldots, n-1$, and $k = n$ because it follows from (4.3)–(4.5) that $\varepsilon_{n-1} = U_2(\rho/2)/\rho U_1(\rho/2) = 1 - \rho^{-2} \leq 0$. In this case the assertion of Theorem 4.1 is obviously correct. On the other hand, if $\rho \geq 2$, then we obtain from (4.6) that $p_{2n-2}^* = 1 - \rho^{-2} \geq \frac{3}{4}$ and by an induction argument that $p_{2j}^* > \frac{1}{2}$, $j = 1, \ldots, n$. In this case we have, from (4.6),

$$(4.7) \qquad p_{2j}^* = 1 - \rho^{-2(n-j)}\left(\prod_{i=j+1}^{n-1} q_{2i}^* p_{2i}^* \right)^{-1}, \qquad j = n-1, \ldots, 1,$$

and it can easily be seen that (4.7) and (4.4) define the same sequences. The

assertion now follows from the representation (4.5) [note that we obtain $k = 1$ in (4.3)]. Finally, if $1 < \rho < 2$ and there exists an index $k - 1$ such that

$$1 - \rho^{-2(n-k+1)}\left(\prod_{i=k}^{n-1} q_{2i}^* p_{2i}^*\right)^{-1} < \tfrac{1}{2},$$

then $p_{2k-2}^* = \tfrac{1}{2}$ and we obtain

$$1 - \rho^{-2(n-k+2)}\left(\prod_{i=k-1}^{n-1} q_{2i}^* p_{2i}^*\right)^{-1} = 1 - \rho^{-2(n-k+1)}\left(\prod_{i=k}^{n-1} q_{2i}^* p_{2i}^*\right)^{-1} 4\rho^{-2} < \tfrac{1}{2},$$

which also implies $p_{2k-4}^* = \tfrac{1}{2}$. Consequently, the sequence of canonical moments of the optimal discriminating design is of the form $(\tfrac{1}{2}, \ldots, \tfrac{1}{2}, p_{2k}^*, \tfrac{1}{2}, \ldots, \tfrac{1}{2}, p_{2n-2}^*, \tfrac{1}{2}, 1)$, where $p_{2k}^*, \ldots, p_{2n-2}^* > \tfrac{1}{2}$ can be calculated recursively by (4.4) [or, equivalently, by (4.7)]. The assertion of the theorem now follows from (4.5), which shows that k is defined by (4.3). $\square$

REMARK 4.2. The proof of Theorem 4.1 shows that there are three different cases for the sequence of canonical moments of the optimal discriminating design ξ^*:

(a) $\rho \leq 1$. ξ^* has canonical moments $(\tfrac{1}{2}, \ldots, \tfrac{1}{2}, 1)$. In this case the optimal discriminating design coincides with the locally maximin designs considered by Spruill (1990) ($b = -a = 1$, $\vartheta = 1$) and the set $\mathcal{N}(\xi^*)$ of Theorem 3.1 is given by $\mathcal{N}(\xi^*) = \{n\}$.

(b) $1 < \rho < 2$. In this case there exists an index $1 \leq k \leq n$ (depending on n and ρ) such that ξ^* has canonical moments

$$\left(\tfrac{1}{2}, \ldots, \tfrac{1}{2}, p_{2k}^*, \tfrac{1}{2}, \ldots, \tfrac{1}{2}, p_{2n-2}^*, \tfrac{1}{2}, 1\right),$$

where $p_{2j}^* > \tfrac{1}{2}$, $j = k, \ldots, n - 1$, is defined recursively by ($p_{2n}^* = 1$)

$$(4.8) \qquad p_{2j}^* = 1 - \frac{\rho^{-2}}{p_{2j+2}^*}, \qquad j = n - 1, \ldots, k,$$

and explicitly given by (4.2). It is straightforward to show that for sufficiently large n this index always satisfies $k > 1$ (this is a consequence of the fact that for $\rho < 2$ the equation $1 - \rho^{-2}/z = z$ has no fixed point and that the case $k = 1$ implies $\tfrac{1}{2} \leq p_{2j}^* < p_{2j+2}^*$, for all $j = 1, \ldots, n - 1$). For this choice we have

$$\mathcal{N}(\xi^*) = \begin{cases} \{k, \ldots, n\}, & \text{if } 2U_{n-k+2}\!\left(\dfrac{\rho}{2}\right) < \rho U_{n-k+1}\!\left(\dfrac{\rho}{2}\right), \\[2mm] \{k-1, k, \ldots, n\}, & \text{if } 2U_{n-k+2}\!\left(\dfrac{\rho}{2}\right) = \rho U_{n-k+1}\!\left(\dfrac{\rho}{2}\right). \end{cases}$$

(c) $\rho \geq 2$. ξ^* has canonical moments $(\tfrac{1}{2}, p_2^*, \tfrac{1}{2}, \ldots, \tfrac{1}{2}, p_{2n-2}^*, \tfrac{1}{2}, 1)$, where the canonical moments of even order $p_{2j}^* > \tfrac{1}{2}$ are defined recursively by (4.8) for all $j = 1, \ldots, n - 1$ [or, equivalently, by (4.2)]. In this case we have $\mathcal{N}(\xi^*) = \{1, \ldots, n\}$.

In the following we will identify the support points and the weights of the optimal discriminating design ξ^* in Theorem 4.1. At first we will consider the case $\rho \geq 2$, which turns out to be essential for the general case.

THEOREM 4.3. *Let $\rho \geq 2$. Then the support points $x_0, \ldots, x_n$ of the optimal discriminating design ξ^* are given by $x_j = z_j + (a + b)/2, j = 0, \ldots, n$, where $z_0, \ldots, z_n$ are the zeros of the polynomial*

$$U_{n+1}\left(\frac{x\vartheta}{2}\right)U_{n-1}\left(\frac{\rho}{2}\right) - U_{n-1}\left(\frac{x\vartheta}{2}\right)U_{n+1}\left(\frac{\rho}{2}\right)$$

and the weights at these points are

$$\xi(\{x_j\}) = \frac{2U_n(z_j\vartheta/2)U_{n-1}(\rho/2)}{U'_{n+1}(z_j\vartheta/2)U_{n-1}(\rho/2) - U'_{n-1}(z_j\vartheta/2)U_{n+1}(\rho/2)},$$

$$j = 0, \ldots, n.$$

PROOF. From Remark 4.2 we have $k = 1$ in Theorem 4.1. The support points and weights of the optimal discriminating design ξ^* are given in Proposition 3.3. It is easy to see that the polynomials $Q_l(x)$ and $P_l(x)$ are functions of $x - (a + b)/2$, and we may assume without loss of generality that the design space is the interval $[-(b - a)/2, (b - a)/2]$, that is, $\tau = (a + b)/2 = 0$ (note that by Theorem 4.1 the canonical moments of the optimal design depend on the interval $[a, b]$ only by its length). By Studden [(1982b), Lemma 2.10], the design $\tilde{\xi}$ with canonical moments $(\frac{1}{2}, \tilde{p}_2, \frac{1}{2}, \ldots, \frac{1}{2}, \tilde{p}_{2n-2}, \frac{1}{2}, 1)$ and

$$(4.9) \qquad \tilde{p}_{2j} = 1 - p^*_{2(n-j)} = 1 - \frac{U_{j+1}(\rho/2)}{\rho U_j(\rho/2)} = \frac{U_{j-1}(\rho/2)}{\rho U_j(\rho/2)},$$

$$j = 1, \ldots, n - 1,$$

has the same support points as ξ^* [here we used (4.2) and the recursive relation for the Chebyshev polynomials of the second kind; see, e.g., Rivlin (1990), page 40]. An expansion of the determinant in (3.7) yields that the support points of $\tilde{\xi}$ are given by zeros of the polynomial $(x^2 - \sigma^2)\tilde{Q}_{n-1}(x)$, where $\sigma = (b - a)/2$ and $\tilde{Q}_l(x)$ is defined recursively by

$$(4.10) \qquad \begin{aligned} \tilde{Q}_{l+1}(x) &= x\tilde{Q}_l(x) - \sigma^2\tilde{p}_{2l}\tilde{q}_{2l+2}\tilde{Q}_{l-1}(x) \\ &= x\tilde{Q}_l(x) - \vartheta^{-2}\frac{U_{l-1}(\rho/2)U_{l+2}(\rho/2)}{U_l(\rho/2)U_{l+1}(\rho/2)}\tilde{Q}_{l-1}(x), \end{aligned}$$

$l = 0, \ldots, n - 2$, $\tilde{Q}_0(x) = 1$, $\tilde{Q}_1(x) = x$. Using the recursive relation for the polynomials $U_l(x)$, it now follows by an induction argument and formula (22.7.26) in Abramowitz and Stegun (1964) that

$$\begin{aligned}(4.11) \quad \tilde{Q}_l(x) &= \frac{U_{l+2}(x\vartheta/2)U_l(\rho/2) - U_l(x\vartheta/2)U_{l+2}(\rho/2)}{\vartheta^l((x\vartheta)^2 - \rho^2)U_l(\rho/2)} \\ &= \frac{U_{l+2}(x\vartheta/2)U_l(\rho/2) - U_l(x\vartheta/2)U_{l+2}(\rho/2)}{\vartheta^{l+2}(x^2 - \sigma^2)U_l(\rho/2)},\end{aligned}$$

and the assertion about the support points is an immediate consequence of Proposition 3.3 (note that $\rho = \vartheta\sigma$). For the second part we remark that the polynomial $Q_{n-1}(x)$ in (3.7) (corresponding to the canonical moments of ξ^*) satisfies $Q_{n-1}(x) = \tilde{Q}_{n-1}(x)$. Because $\rho \geq 2$ we have $\mathcal{N}(\xi^*) = \{1, \ldots, n\}$ and obtain, from (4.1) and the discussion in the proof of Theorem 4.1,

$$(4.12) \qquad q_{2l}^* p_{2l+2}^* = \rho^{-2} = \sigma^{-2}\vartheta^{-2}, \qquad l = 1, \ldots, n-1.$$

For the polynomials $P_l(x)$ defined in (3.6), this implies that

$$P_{l+1}(x) = xP_l(x) - \vartheta^{-2}P_{l-1}(x), \qquad l = 0, 1, \ldots, n-1,$$

$P_0(x) = 1$ and $P_1(x) = x$, and it is easy to see that these polynomials are given by

$$P_l(x) = \vartheta^{-l}U_l\left(\frac{x\vartheta}{2}\right).$$

The assertion about the weights now follows directly from Proposition 3.3 and (4.11). □

THEOREM 4.4. *Let* $\rho > 0$ *and let* k *be defined by* (4.3). *Then the support points* $x_0, \ldots, x_n$ *of the optimal discriminating design are given by* $x_j = z_j + (a+b)/2$, $j = 0, \ldots, n$, *where* $z_0, \ldots, z_n$ *are the zeros of the polynomial*

$$\begin{aligned}H_{n+1}(x) = U_{k-1}&\left(\frac{x\vartheta}{\rho}\right)\left\{U_{n-k+2}\left(\frac{x\vartheta}{2}\right)U_{n-k}\left(\frac{\rho}{2}\right) - U_{n-k}\left(\frac{x\vartheta}{2}\right)U_{n-k+2}\left(\frac{\rho}{2}\right)\right\} \\ &- U_{k-2}\left(\frac{x\vartheta}{\rho}\right)\left\{U_{n-k+1}\left(\frac{x\vartheta}{2}\right)U_{n-k-1}\left(\frac{\rho}{2}\right)\right. \\ &\left. - U_{n-k-1}\left(\frac{x\vartheta}{2}\right)U_{n-k+1}\left(\frac{\rho}{2}\right)\right\},\end{aligned}$$

while the weights are given by

$$\begin{aligned}&\xi(\{x_j\}) \\ &= \frac{U_{n-k+1}(z_j\vartheta/2)U_{k-1}(z_j\vartheta/\rho)U_{n-k}(\rho/2) - U_{n-k}(z_j\vartheta/2)U_{k-2}(z_j\vartheta/\rho)U_{n-k+1}(\rho/2)}{\vartheta^{-1}\cdot(d/dx)H_{n+1}(x)|_{x=z_j}},\end{aligned}$$

$j = 0, \ldots, n$.

PROOF. As in the proof of Theorem 4.3 we assume without loss of generality that the design space is given by the interval $[-(b-a)/2, (b-a)/2]$ and consider the reversed sequence $(\frac{1}{2}, \tilde{p}_2, \frac{1}{2}, \ldots, \frac{1}{2}, \tilde{p}_{2n-2}, \frac{1}{2}, 1)$, where $\tilde{p}_{2j} = 1 - p^*_{2(n-j)}$, $j = 1, \ldots, n-1$. The corresponding design $\tilde{\xi}$ has the same support points as ξ^*, and from the definition of k in (4.3) and Remark 4.2 we have $\tilde{p}_{2j} = \frac{1}{2}$, $j = n-k+1, \ldots, n-1$. Thus it follows from (3.7) that $[\sigma = (b-a)/2]$

$$
\tilde{Q}_{n-1}(x) = K \left(\begin{array}{ccccccc} \overbrace{-\left(\frac{\sigma}{2}\right)^2 \cdots -\left(\frac{\sigma}{2}\right)^2}^{k-2} & -\sigma^2 \tilde{p}_{2(n-k)} \tilde{q}_{2(n-k+1)} & \cdots & -\sigma^2 \tilde{p}_2 \tilde{q}_4 \\ x \quad x \quad \cdots \quad x & x & \cdots & x \quad x \end{array} \right)
$$

$$
= K \left(\begin{array}{c} \overbrace{-\left(\frac{\sigma}{2}\right)^2 \cdots -\left(\frac{\sigma}{2}\right)^2}^{k-2} \\ x \quad x \quad \cdots \quad x \end{array} \right) K \left(\begin{array}{cccc} -\sigma^2 \tilde{p}_{2(n-k-1)} \tilde{q}_{2(n-k)} & \cdots & -\sigma^2 \tilde{p}_2 \tilde{q}_4 \\ x & & x \quad \cdots \quad x & x \end{array} \right)
$$

$$
- \sigma^2 \tilde{p}_{2(n-k)} \tilde{q}_{2(n-k+1)} K \left(\begin{array}{c} \overbrace{-\left(\frac{\sigma}{2}\right)^2 \cdots -\left(\frac{\sigma}{2}\right)^2}^{k-3} \\ x \quad x \quad \cdots \quad x \quad x \end{array} \right)
$$

$$
\times K \left(\begin{array}{ccc} -\sigma^2 \tilde{p}_{2(n-k-2)} \tilde{q}_{2(n-k-1)} & & -\sigma^2 \tilde{p}_2 \tilde{q}_4 \\ x & & x \quad \cdots \quad x \quad x \end{array} \right),
$$

where the last line follows from Sylvester's identity [see Lau (1983), page 23]. Observing the reasoning (4.10) and (4.11) in the proof of Theorem 4.3, we see that the second factors in both terms are given by (4.11) for $l = n-k$ and $l = n-k-1$, respectively. By the same reasoning as at the end of the proof of Theorem 4.3, the first factors satisfy the recurrence relation

$$
\tilde{P}_{l+1}(x) = x\tilde{P}_l(x) - \left(\frac{\sigma}{2}\right)^2 \tilde{P}_{l-1}(x),
$$

$\tilde{P}_{-1}(x) = 0$ and $\tilde{P}_0(x) = 1$ and are given by $(\sigma/2)^{k-1} U_{k-1}(x/\sigma)$ and $(\sigma/2)^{k-2} U_{k-2}(x/\sigma)$, respectively. Thus it follows from (4.2) and $\tilde{p}_{2j} = 1 - p^*_{2(n-j)}$ that

$$
\tilde{Q}_{n-1}(x) = \left(\frac{\sigma}{2}\right)^{k-1} \left[U_{k-1}\left(\frac{x\vartheta}{\rho}\right) Q^*_{n-k}(x) \right.
$$
$$
\left. - \frac{\sigma}{\rho} \frac{U_{n-k-1}(\rho/2)}{U_{n-k}(\rho/2)} U_{k-2}\left(\frac{x\vartheta}{\rho}\right) Q^*_{n-k-1}(x) \right],
$$

where $Q_l^*(x)$ is defined by (4.11). This yields

$$(4.13) \qquad (x^2 - \sigma^2)\tilde{Q}_{n-1}(x) = \left(\frac{\sigma}{2}\right)^{k-1} \vartheta^{k-n-2} \cdot \frac{H_{n+1}(x)}{U_{n-k}(\rho/2)}$$

and proves the assertion for the support points. For the calculation of the weights of the design ξ^* we obtain for the polynomial $P_n(x)$ in (3.6), by a similar reasoning,

$$\begin{aligned}
P_n(x) = \frac{(\sigma/2)^{k-1} \vartheta^{k-n-1}}{U_{n-k}(\rho/2)} &\left\{ U_{n-k+1}\left(\frac{x\vartheta}{2}\right) U_{k-1}\left(\frac{x\vartheta}{\rho}\right) U_{n-k}\left(\frac{\rho}{2}\right) \right. \\
(4.14) & \\
&\left. - U_{n-k}\left(\frac{x\vartheta}{2}\right) U_{k-2}\left(\frac{x\vartheta}{\rho}\right) U_{n-k+1}\left(\frac{\rho}{2}\right) \right\},
\end{aligned}$$

and Theorem 4.4 follows by a further application of Proposition 3.3. $\square$

REMARK 4.5. If $[a, b] = [-1, 1]$, $\vartheta = 1$, the optimal discriminating design coincides with the locally maximin design determined by Spruill (1990). In this case we have $\rho = 1$ and $k = n$ in Theorem 4.1, and the polynomial $H_{n+1}(x)$ in Theorem 4.4 is given by

$$U_{n-1}(x)(x^2 - 1).$$

Moreover, it is not too hard to show that the weights at the support points $x_j = -\cos((\pi/n)j)$, $j = 0, \ldots, n$, are proportional to $1 : 2 : \cdots : 2 : 1$ [see Spruill (1990)]. Another case of interest is $\rho = 2$ because at this point the sequence in (4.4) changes from a divergent sequence into a convergent sequence ($j \to -\infty$). Observing (4.2) and $U_n(1) = n + 1$, we obtain

$$p_{2j}^* = \frac{n - j + 2}{2(n - j) + 2}, \qquad j = 1, \ldots, n,$$

$p_{2j-1}^* = \frac{1}{2}$, and Theorem 2.5 of Dette (1992) yields that in this case the optimal discriminating design ξ^* has masses proportional to $2 : 3 : \cdots : 3 : 2$ at the zeros of the polynomial

$$(x^2 - 1)U_n'(x).$$

5. Asymptotic distributions. In this section we consider the same setting as in Section 4 when the degree n of the polynomial is large. By Theorem 4.1 and (4.8), the canonical moments (of even order) of the optimal discriminating design ξ^* are given by

$$(5.1) \qquad p_{2j}^* = \begin{cases} \dfrac{1}{2}, & 1 \le j \le k - 1, \\[3mm] 1 - \dfrac{\rho^{-2}}{p_{2j+2}^*}, & k \le j \le n - 1, \end{cases}$$

where k is defined in (4.3). If $\rho < 2$, then it follows from the discussion in Remark 4.2 that for sufficiently large n there exists an index j_0 such that

$p_{2i}^* = \frac{1}{2}$ for all $i \le n - j_0$. Consequently, the optimal discriminating design converges (weakly) to the probability measure with canonical moments $p_i = \frac{1}{2}$ for all $i \in \mathbb{N}$. If $\rho \ge 2$, then Remark 4.2 shows that $p_{2j}^* > \frac{1}{2}$ for all $j \in \mathbb{N}$. The sequence (5.1) is decreasing (i.e., $p_{2j-2}^* < p_{2j}^*$) and consequently, as n tends to infinity, the canonical moments of the optimal design ξ^* converge to the distribution for which

$$p_{2j} = z, \qquad p_{2j-1} = \tfrac{1}{2},$$

for all $j \in \mathbb{N}$, where $z \ge \frac{1}{2}$ is the fixed point of the equation $z = 1 - \rho^{-2}/z$, that is,

$$(5.2) \qquad z = \tfrac{1}{2} + \tfrac{1}{2}\sqrt{1 - 4\rho^{-2}}.$$

THEOREM 5.1.

(a) If $\rho < 2$, then the optimal discriminating design ξ^* converges weakly to the arcsin distribution with density

$$f(x) = \begin{cases} \dfrac{1}{\pi} \dfrac{1}{\sqrt{(x-a)(b-x)}}, & \text{if } a < x < b, \\ 0, & \text{otherwise.} \end{cases}$$

(b) If $\rho \ge 2$, then the optimal discriminating design ξ^* converges weakly to a distribution η which has a discrete and an absolute continuous component. The absolute continuous component of the limit distribution η is given by

$$f(x) = \begin{cases} \dfrac{2(b-a)}{\pi} \dfrac{\sqrt{16\vartheta^{-2} - (2x - (a+b))^2}}{\left[(b-a) + \sqrt{(b-a)^2 - 16\vartheta^{-2}}\right]\left[(b-a)^2 - (2x - (a+b))^2\right]}, \\ \qquad \text{if } \left|x - \dfrac{a+b}{2}\right| < 2|\vartheta|^{-1}, \\ 0, \qquad \text{otherwise,} \end{cases}$$

and, additionally, η has two isolated jumps of magnitude

$$\frac{4 - \rho^2 + \rho\sqrt{\rho^2 - 4}}{4}$$

at the points a and b.

PROOF. By the discussion at the beginning of this section, we have to identify the (unique) probability measure η corresponding to the sequence of canonical moments

$$\tfrac{1}{2}, z, \tfrac{1}{2}, z, \tfrac{1}{2}, z, \ldots,$$

where $z \geq \frac{1}{2}$. Let $\sigma = (b - a)/2$ and $\tau = (a + b)/2$. Then it follows [see, e.g., Dette (1992), page 245] that the Stieltjes transform of η is given by the continued fraction expansion

$$\Phi(w) = \int_a^b \frac{d\eta(t)}{w - t} = \frac{1}{w - \tau - \sigma^2 z G(w)},$$

where

$$G(w) = \frac{1}{|w - \tau|} - \frac{\sigma^2 z(1 - z)|}{|\quad w - \tau} - \frac{\sigma^2 z(1 - z)|}{|\quad w - \tau} - \cdots$$

$$= \frac{1}{w - \tau - \sigma^2 z(1 - z) G(w)}.$$

Solving with respect to $G(w)$ yields

$$G(w) = \frac{w - \tau - \sqrt{(w - \tau)^2 - 4\sigma^2(1 - z)z}}{2\sigma^2 z(1 - z)},$$

where the square root is defined such that

(5.3)
$$\left| \frac{w - \tau}{2\sigma\sqrt{z(1 - z)}} + \sqrt{\frac{(w - \tau)^2}{4\sigma^2 z(1 - z)} - 1} \right| > 1$$

[see VanAssche, (1987), page 176], and consequently we obtain, by straightforward calculation,

$$\Phi(w) = -\frac{1}{2z} \frac{(1 - 2z)(w - \tau) - \sqrt{(w - \tau)^2 - 4\sigma^2 z(1 - z)}}{(w - \tau)^2 - \sigma^2}.$$

A delicate analysis of the sign of the square root defined by (5.3) shows that Φ has no poles if $z \leq \frac{1}{2}$. If $z > \frac{1}{2}$ the Stieltjes transform of η has two simple poles at the points a and b with residues

(5.4)
$$\frac{2z - 1}{2z} = \frac{4 - \rho^2 + \rho\sqrt{\rho^2 - 4}}{4}.$$

From the inversion formula for Stieltjes transforms [see, e.g., Perron (1954)], it now follows that the absolute continuous component of the limit distribution η is given by

(5.5)
$$f(x) = \frac{1}{2\pi z} \frac{\sqrt{4\sigma^2 z(1 - z) - (x - \tau)^2}}{\sigma^2 - (x - \tau)^2},$$

where $|x - \tau|^2 \leq 4\sigma^2 z(1 - z)$. If $\rho \geq 2$, then $4\sigma^2 z(1 - z) = 4\vartheta^{-2}$ [which follows from (4.12)], $\tau = (a + b)/2$ and, using (5.2), the absolute continuous component of η can be written as

$$f(x) = \frac{2(b - a)}{\pi}$$

$$\times \frac{\sqrt{16\vartheta^{-2} - (2x - (a + b))^2}}{\left[(b - a) + \sqrt{(b - a)^2 - 16\vartheta^{-2}}\right]\left[(b - a)^2 - (2x - (a + b))^2\right]},$$

$|x - (a + b)/2| \leq 2|\vartheta|^{-1}$. Additionally, the limit distribution η has two jumps of magnitude $(2z - 1)/2z$ at the points a and b, which proves part (b) of the theorem.

Finally, if $\rho < 2$, it follows that $z = \frac{1}{2}$ and η is absolutely continuous. For the density, we obtain from (5.5), by straightforward calculations,

$$f(x) = \frac{1}{\pi} \frac{1}{\sqrt{(b - x)(x - a)}}, \qquad a < x < b,$$

which proves part (a) of the theorem. $\square$

REMARK 5.2. If $\rho \leq 2$, then the limit distribution η of the optimal discriminating design is the arcsin law, which is also the limit distribution of many other sequences of designs [see, e.g., Lau (1988]. If $\rho \geq 2$, the situation is changing and η has two additional jumps at the points a and b. For example, if $[a, b] = [-1, 1]$ and $\vartheta = 2.5 = \rho$, we obtain that η has masses $\frac{3}{8}$ at the points -1 and 1, while the remaining mass is distributed over the interval $[-\frac{4}{5}, \frac{4}{5}]$ according to the "density"

$$f(x) = \frac{\sqrt{16 - 25x^2}}{8\pi(1 - x^2)} I\left\{|x| \leq \frac{4}{5}\right\}.$$

Note that in the case $\rho \to \infty$ we have $z \to 1$ [from (5.2)] and (5.4) yields that the limit distribution approximates the two-point design with equal masses at the points a and b.

EXAMPLE 5.3. It might be of interest how the designs behave in a situation where they are not optimal. To this end consider the case $a = -b$, $\vartheta = 1$, $\rho = b$ and $n = 3$ (cubic regression). If $b \leq \sqrt{2}$, then the locally maximin design determined by Spruill (1990) is also the optimal discriminating design, that is,

$$\xi_1 = \begin{pmatrix} -b & -\dfrac{b}{2} & \dfrac{b}{2} & b \\[2mm] \dfrac{1}{6} & \dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{6} \end{pmatrix}$$

and the canonical moments (of even order) of ξ_1 are $p_2 = p_4 = \frac{1}{2}$ and $p_6 = 1$ (this follows from Theorems 4.1 and 4.3). If $\sqrt{2} \le b \le \sqrt{3}$, we obtain from the results of Section 4 that the optimal discriminating design is given by

$$
\xi_2 = \begin{pmatrix} -b & -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} & b \\ \dfrac{1}{2}\dfrac{b^2-1}{2b^2-1} & \dfrac{1}{2}\dfrac{b^2}{2b^2-1} & \dfrac{1}{2}\dfrac{b^2}{2b^2-1} & \dfrac{1}{2}\dfrac{b^2-1}{2b^2-1} \end{pmatrix},
$$

with canonical moments (of even order) $p_2 = \frac{1}{2}$, $p_4 = 1 - b^{-2}$ and $p_6 = 1$. Note that the interior support points are independent of $b \in [\sqrt{2}, \sqrt{3}]$. Finally, if $b \ge \sqrt{3}$, the optimal design is

$$
\xi_3 = \begin{pmatrix} -b & -\sqrt{\dfrac{b^2-2}{b^2-1}} & \sqrt{\dfrac{b^2-2}{b^2-1}} & b \\ \dfrac{1}{2}\left(1 - \dfrac{b^2}{b^4-2b^2+2}\right) & \dfrac{1}{2}\dfrac{b^2}{b^4-2b^2+2} & \dfrac{1}{2}\dfrac{b^2}{b^4-2b^2+2} & \dfrac{1}{2}\left(1 - \dfrac{b^2}{b^4-2b^2+2}\right) \end{pmatrix},
$$

with canonical moment (of even order) $p_2 = (b^2-2)/(b^2-1)$, $p_4 = 1 - b^{-2}$ and $p_6 = 1$ (this follows from Theorem 4.3 or 4.4). For the values of the optimality criterion (4.1) at these points, we obtain the following:

$$
\Phi(\xi_1) = \begin{cases} \dfrac{b^6}{16}, & \text{if } b \le \sqrt{2}, \\[2mm] \dfrac{b^4}{8}, & \text{if } \sqrt{2} \le b \le 2, \\[2mm] \dfrac{b^2}{2}, & \text{if } 2 \le b; \end{cases}
$$

$$
\Phi(\xi_2) = \begin{cases} \dfrac{b^2}{4}(b^2-1), & \text{if } 1 < b \le \sqrt{3}, \\[2mm] \dfrac{b^2}{2}, & \text{if } b > \sqrt{3}; \end{cases}
$$

$$
\Phi(\xi_3) = b^2 \dfrac{b^2-2}{b^2-1} \quad \text{for all } b \ge \sqrt{2}.
$$

Table 1 gives some efficiencies for different values of the parameter b. Note that the designs ξ_2 and ξ_3 are not defined if b is too small (i.e., $b \le 1$ for ξ_2 and $b \le \sqrt{2}$ for ξ_3). This is caused by the fact that the designs have to satisfy $\mathcal{M}(\xi_2) = \{2,3\}$ and $\mathcal{M}(\xi_3) = \{1,2,3\}$, which is impossible if b is too small. We conclude with the statement that the performance of the different designs

TABLE 1
Efficiencies of the designs ξ_1, ξ_2 and ξ_3 for different intervals $[-b, b]$

					b				
Design	$\sqrt{2}$	1.5	1.6	1.7	$\sqrt{3}$	2	3	10	∞
ξ_1	1	0.9	0.8205	0.7646	0.75	0.75	0.5714	0.5051	0.5
ξ_2	1	1	1	1	1	0.75	0.5714	0.5051	0.5
ξ_3	0	0.64	0.9204	0.9966	1	1	1	1	1

with respect to the criterion (2.1) will depend heavily on the length of the interval.

REFERENCES

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

ANDERSON, T. W. (1962). The choice of the degree of a polynomial regression as a multiple decision problem. *Ann. Math. Statist.* **33** 255–265.

CHANG, F. C., KEMPERMAN, J. B. and STUDDEN, W. J. (1993). A normal limit theorem for moment sequences. *Ann. Probab.* **21** 1295–1309.

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

DETTE, H. (1992). Optimal designs for a class of polynomials of odd and even degree. *Ann. Statist.* **20** 238–259.

DETTE, H. (1993). Elfving's theorem for *D*-optimality. *Ann. Statist.* **21** 753–766.

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 and II. Technical Reports 83-23, 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.

PERRON, O. (1954). *Die Lehre von den Kettenbrüchen* **1, 2**. Teubner, Stuttgart.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

RIVLIN, T. J. (1990). *Chebyshev Polynomials.* Wiley, New York.

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

SPRUILL, M. G. (1990). Good designs for testing the degree of a polynomial mean. *Sankhyā Ser. B* **52** 67–74.

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. Gupta and J. O. Berger, eds.) 335–350. Academic Press, New York.

VANASSCHE, W. (1987). *Asymptotics for Orthogonal Polynomials. Lecture Notes in Math.* **1265**. Springer, Berlin.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
ABTEILUNG MATHEMATIK
TECHNISCHE UNIVERSITÄT DRESDEN
MOMMSENSTRASSE 13
01062 DRESDEN
GERMANY