OPTIMAL EXACT DESIGNS FOR POLYNOMIAL REGRESSION

By

Kenneth B. Constantine[1]

and

W. J. Studden[2]

Mimeograph Series #81-35

[1]Research supported by a NSF Graduate Fellowship and National Science Foundation. Grant number MCS-7901707.

[2]Research supported by National Science Foundation. Grant number MCS-7901707.

1. Introduction

Consider the linear model $y(x) = \theta' f(x) + \varepsilon$ which is assumed to hold for each "level" $x \in \mathcal{X}$ (compact). Here ε denotes a random variable with mean 0 and variance σ^2 for all x. The present interest is in (univariate) polynomial regression of degree n on $\mathcal{X} = [a,b]$ so that $\theta = (\theta_0, \theta_1, \ldots, \theta_n)'$ and $f(x) = (1, x, \ldots, x^n)'$.

Suppose that N uncorrelated observations on the response $y(x)$ are to be obtained at levels $x_1, \ldots, x_N$. The linear model for these data is $Y = X\theta + e$, where $Y = [y(x_1), \ldots, y(x_N)]'$, where $X_{ij} = f_j(x_i)$ for $1 \le i \le N$ and $0 \le j \le n$, and where $e = (\varepsilon_1, \ldots, \varepsilon_N)'$. It will be assumed that inferences about θ are to be based on the classical estimator $\hat{\theta} = (X'X)^{-1}X'Y$. Thus $E(\hat{\theta}) = \theta$ and $\mathrm{Cov}(\hat{\theta}) = \sigma^2(X'X)^{-1}$. Note that if rank $(X) < n+1$, then the inverse operation should be interpreted as a generalized inverse. The design goal, which will be made more precise, is to choose $x_1, \ldots, x_N$ so as to "minimize" $(X'X)^{-1}$.

In order to more conveniently formulate the design problem, let $x_0, \ldots x_r$ now denote the distinct levels at which $n_0, \ldots, n_r$ observations are taken. Here $n_0 + \ldots + n_r = N$. An "exact design" ξ^N is a probability measure on $\mathcal{X}$ which concentrates mass n_i/N at each x_i. Such a design prescribes exactly where and how to allocate observations. The set of all exact designs for a given value of N will be denoted by Ξ_N. The "information matrix (per observation)" of an exact design ξ^N is $M(\xi^N) = \int_{\mathcal{X}} f(x)f(x)' d\xi^N(x)$. It is readily shown that $\mathrm{Cov}(\hat{\theta}) = \sigma^2 M^{-1}(\xi^N)/N$. Thus a reformulation of the design problem is to determine an exact design ξ^N which "minimizes" $M^{-1}(\xi^N)$. Note also that for polynomial regression, $M_{ij}(\xi^N) = \mu_{i+j}$ where $0 \le i,j \le n$ and each $\mu_k = \int_a^b x^k d\xi^N(x) = \sum_{\ell=0}^{r} n_\ell x_\ell^k / N$.

An approach which is often taken in optimal design work is to extend consideration to the class of all "approximate designs", ie. arbitrary probability measures ξ on $\mathcal{X}$. This approach has the distinct advantage of greater mathematical tractability. Its limitation is that, in practice, only an exact design may be implemented. It is often the case that an optimal approximate design is not exact for certain choices of N (or even for any choice of N). This limitation will be especially important when N is not too large.

The present interest is to address some classical optimal design questions in the context of the exact design setting. The results obtained (and conjectured) will be compared with results known for approximate designs.

Section 2 is denoted to the admissibility problem for polynomial regression. Theorem 2.1 provides a necessary condition for admissibility. It is conjectured that this condition is also sufficient and the basis for the conjecture is discussed.

Section 3 treats the design criterion of D-optimality. Salaevskii (1966) conjectures that a D-optimal exact design ξ_0^N distributes observations as evenly as possible among the n+1 support points of the D-optimal approximate design. Theorem 3.1 provides a simplified proof of Salaevskii's result that the conjecture holds for sufficiently large N.

Section 4 provides some examples of G-optimal exact designs.

2. Admissibility

Recall that the exact design problem is to determine an exact design ξ^N which "minimizes" $M^{-1}(\xi^N)$. A particular optimality criterion may correspond to a real-valued function Φ on the set of non-negative definite matrices. A "Φ-optimal" design would minimize $\Phi(M^{-1}(\xi^N))$ among all exact designs. The examples of $\Phi(M^{-1}(\xi^N))=|M^{-1}(\xi^N)|$ ("D-optimality") and of $\Phi(M^{-1}(\xi^N))=\max_{\mathcal{X}} f(x)'M^{-1}(\xi^N)f(x)$ ("G-optimality") will be considered in sections 3 and 4.

In many cases (including D and G-optimality) the function Φ is monotone in the sense that if $M^{-1}(\xi_1^N) \leq M^{-1}(\xi_2^N)$, then $\Phi(M^{-1}(\xi_1^N)) \leq \Phi(M^{-1}(\xi_2^N))$. Here the inequality $A \leq B$ for non-negative definite matrices A and B should have the customary meaning that B-A is non-negative definite. These developments naturally suggest the admissibility problem: characterize those exact designs whose inverse information matrices are minimal with respect to "$\leq$". Equivalently, the problem is to characterize the exact designs whose information matrices are maximal with respect to "$\leq$". Accordingly an exact design ξ^N is "admissible" if and only if there exists no other exact design $\tilde{\xi}^N$ such that $M(\tilde{\xi}^N) \geq M(\xi^N)$.

In the case of polynomial regression, the following lemma relates the admissibility problem to a problem involving the moments $\mu_1,\ldots,\mu_{2n-1}$, μ_{2n}.

<u>Lemma 2.1:</u> ξ^N is admissible for polynomial regression of degree n if and only if there exists no other exact design which shares the same values of $\mu_1,\ldots,\mu_{2n-1}$ but has a larger value of μ_{2n}.

<u>Proof:</u> See Karlin and Studden (1966).

In the approximate setting, a design ξ is admissible for polynomial regression of degree n if and only if the support of ξ includes n-1 or fewer interior points. This characterization has been developed by de la Garza (1954), Kiefer (1959), and Karlin and Studden (1966). Note that this admissibility condition involves only the support of an approximate design. It will be seen that the corresponding statement for exact designs does not involve only the support of an exact design.

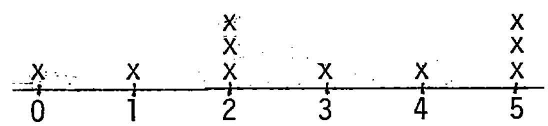
The following definition establishes some terminology which will be used in subsequent developments.

Definition 2.1: i. If $\xi^N(\{x_j\}) > 1/N$, then x_j is termed a cluster of ξ^N.

ii. If $\xi^N(\{x_j\}) = 1/N$, then x_j is termed a singlet of ξ^N.

Example 2.1: Let [a,b] = [0,5], let N=10, and let $\xi^N = .1\delta_0 + .1\delta_1 + .3\delta_2 + .1\delta_3 + .1\delta_4 + .3\delta_5$. (Here δ_x denotes a point mass at x.)

ξ^N:



Thus ξ^N is comprised of an interior cluster 2, interior singlets 1,3,& 4, a cluster 5, and a singlet 0.

It is proposed that the clusters of an exact design correspond to the support points of an approximate design. The twist to this relationship is that pairs of adjacent interior singlets and singlets at a or b act as clusters. In this spirit, the following theorem establishes conditions that are necessary for an exact design to be admissible for polynomial regression of degree n. Arguments for the sufficiency of these conditions will be given after the proof of their necessity.

Theorem 2.1: Let ξ^N be an exact design with r interior support points, with m interior clusters, and with s pairs of adjacent interior singlets. If ξ^N is admissible for polynomial regression of degree n, then:

i. $r \leq 2n-1$,

ii. $m \leq n-1$, and

iii. $s \leq (n-1) - m$.

Proof. The proof will make repeated use of polynomials of the form

$$P(x) = \prod_{j=1}^{2n} (x-y_j) = \sum_{\ell=0}^{2n} (-1)^{\ell} e_j x^{2n-\ell}, \tag{2.1}$$

where $a \leq y_1,\ldots,y_{2n} \leq b$ and where each

$$e_k = \sum_{1 \leq j_1 < \ldots < j_k \leq n} \prod_{\ell=1}^{k} y_{j_\ell}.$$

Here $e_0=1$. It will be convenient to define $s_k = \sum_{j=1}^{2n} y_j^k$ for $k=0,\ldots,2n$. In terms of this notation,

$$s_k = \sum_{\ell=1}^{k} (-1)^{\ell} e_\ell s_{k-\ell}$$

for $k=1,\ldots,2n$. These equations establish a 1-1 correspondence between $s_1,\ldots,s_k$ and $e_1,\ldots,e_k$ for each $k=1,\ldots,2n$. Furthermore, it is seen that another set of points $\tilde{y}_1,\ldots,\tilde{y}_{2n}$ achieves $\tilde{s}_k = s_k$ for $k=1,\ldots,2n-1$ but $\tilde{s}_{2n} > s_{2n}$ if and only if $\tilde{e}_k = e_k$ for $k=1,\ldots,2n-1$ but $\tilde{e}_{2n} < e_{2n}$. According to (2.1), this is possible if and only if there exists $\varepsilon > 0$ such that the polynomial $\tilde{P}(x) = P(x) - \varepsilon$ has 2n roots on $[a,b]$. (Here $\varepsilon = e_{2n} - \tilde{e}_{2n}$ and the roots of $\tilde{P}$ are $\tilde{y}_1,\ldots,\tilde{y}_{2n}$.) This approach will now be applied to the admissibility problem by appropriate choice of $y_1,\ldots,y_{2n}$.

To demonstrate that conditions i. - iii. must hold, suppose first that an exact design ξ^N has more than $2n-1$ interior support points. Let $y_1,\ldots,y_{2n}$ denote 2n of them. Then it is clear that there exists $\varepsilon > 0$ such that $\tilde{P}(x) = P(x) - \varepsilon$ has roots $\tilde{y}_1,\ldots,\tilde{y}_{2n}$ on $[a,b]$. According to the preliminary argument, this implies that $\tilde{s}_k = s_k$ for $k=1,\ldots,2n-1$ but $\tilde{s}_{2n} > s_{2n}$. Now let $\tilde{\xi}^N$ be the exact design obtained from ξ^N by exchanging the observations at $y_1,\ldots,y_{2n}$ for observations at $\tilde{y}_1,\ldots,\tilde{y}_{2n}$. Then $\tilde{\mu}_k = \mu_k$ for $k=1,\ldots,2n-1$ but $\tilde{\mu}_{2n} > \mu_{2n}$. That is, according to lemma 2.1, ξ^N is inadmissible. Therefore, an admissible design can have no more than 2n-1 interior support points.

Suppose next that ξ^N has more than n-1 interior clusters. Let $x_1,\ldots,x_n$ denote n of them and let $y_{2i-1} = y_{2i} = x_i$ for $i=1,\ldots,n$. As in the previous case, the polynomial P(x) may be lowered to yield alternate observation points $\tilde{y}_1,\ldots,\tilde{y}_{2n}$ such that $\tilde{\mu}_k = \mu_k$ for $k=1,\ldots,2n-1$ but $\tilde{\mu}_{2n} > \mu_{2n}$. Thus ξ^N is inadmissible, implying that an admissible exact design can have no more than n-1 interior clusters.

Suppose finally that ξ^N has $m \leq n-1$ interior clusters and more than (n-1)-m pairs of adjacent interior singlets. Then let $x_1,\ldots,x_m$ denote the interior clusters, let $y_{2i-1} = y_{2i} = x_i$ for $i=1,\ldots,m$, and let $y_{2m+1},\ldots,y_{2n}$ denote points which comprise pairs of adjacent interior singlets. By applying the same method to construct $\tilde{\xi}^N$, it is seen that ξ^N is inadmissible. Therefore, an admissible exact design with $m \leq n-1$ interior clusters can have no more than (n-1)-m pairs of adjacent interior singlets and the proof is complete.

Theorem 2.1 provides a complete class of exact designs for polynomial regression. A "typical" exact design from this class might have clusters at a & b, n-1 clusters with (a,b), and n singlets separating the clusters. A "less typical" exact design from this class, for $n \geq 3$, is given by Example 2.1.

At this time, the sufficiency of the conditions of Theorem 2.1 may only be conjectured.

<u>Conjecture 2.1</u>: If an exact design ξ^N satisfies conditions i.-iii. of Theorem 2.1, then it is admissible for polynomial regression of degree n.

It is believed that this conjecture is valid because if an exact design satisfies the conditions i.-iii., then no other exact design which also satisfies them can achieve the same values of $\mu_1,\ldots,\mu_{2n-1}$. If true in general, this uniqueness property (in addition to lemma 2.1) would prove the conjecture.

In the special case of linear regression, the validity of Conjecture 2.1 is readily demonstrated. Theorem 2.1 implies that an admissible exact design must have the form $\xi^N = (n_0\delta_a + \delta_x + n_1\delta_b)/N$, where $n_0+n_1+1=N$ and $a \le x \le b$. It is clear that no other exact design of this form can achieve $\mu_1=(n_0a + x + n_1b)/N$.

3. D-Optimality

As already remarked, a D-optimal exact design ξ_0^N minimizes $|M^{-1}(\xi^N)|$. Equivalently, $|M(\xi_0^N)| = \max_{\xi^N\in\Xi_N} |M(\xi^N)|$.

Hoel (1958) has obtained the result that an approximate design is D-optimal for polynomial regression of degree n on [a,b] = [-1,1] if and only if it concentrates equal mass at the roots of $\pi(x) = (1-x^2)T_n'(x)$, where $T_n(x)$ is the n^{th} Legendre polynomial. For purposes of notation, let $-1=x_0^0 < x_1^0 <\ldots<x_n^0 = 1$ denote the roots of $\pi(x)$ and let ξ_0 denote the D-optimal approximate design.

If N is an integer multiple of n+1, then the D-optimal exact design coincides with the D-optimal approximate design. Otherwise, a reasonable exact design might be one which distributes the N observations as evenly as possible among the same points $\{x_0^0,\ldots,x_n^0\}$. That such a property characterizes the D-optimal exact design(s) is the subject of the following conjecture of Salaevskii (1966).

<u>Conjecture 3.1</u>: An exact design ξ_*^N is D-optimal for polynomial regression of degree n on [-1,1] if and only if Suppport $(\xi_*^N) = \{x_0^0,\ldots,x_n^0\}$ and $|\xi_*^N(\{x_i^0\}) - \xi_*^N(\{x_j^0\})| \leq 1/N$ whenever $0 \leq i < j \leq n$.

It should be noted that ξ_*^N is unique if and only if N is an integer multiple of n+1 (in which case $\xi_*^N = \xi_0$).

A general proof of Conjecture 3.1 cannot be given at this time. It may be noted that, in order to prove the conjecture, it need only be shown that ξ_0^N has no more than n+1 support points. (It must have at least that many if $|M(\xi_0^N)| > 0$.) Then $|M(\xi_0^N)| = \prod_{i=0}^{n} (n_i^N/N)\, V^2(x_0^N,\ldots,x_n^N)$,

where $V^2(x_0^N,\ldots,x_n^N) = \prod_{0\leq i<j\leq n} (x_j^N - x_i^N)^2$ is the square of the Vandermonde determinant corresponding to the points $x_0^N,\ldots x_n^N$. This quantity is maximized if and only if $x_i^N = x_i^0$ for i=0,...,n. Also, the product $\prod_{i=0}^{n}(n_i^N/N)$ is maximized if and only if $|n_i^N - n_j^N| \leq 1$ whenever $0 \leq i < j \leq n$.

The main result of this section is Theorem 3.1 which states that Conjecture 3.1 holds for large enough N. The proof of the theorem follows (but streamlines) that of Salaevskii (1966). Special cases of Conjecture 3.1 and numerical work which support the validity of the conjecture will be presented following the proof of theorem 3.1.

The following notation will be used extensively. Let

$$\psi_N \begin{bmatrix} x_0,\ldots,x_r \\ n_0,\ldots,n_r \end{bmatrix} = N^{n+1}\ |M(\xi^N)|.$$

According to the Binet-Cauchy formula,

$$\psi_N \begin{bmatrix} x_0,\ldots,x_r \\ n_0,\ldots,n_r \end{bmatrix} = \sum_{1\le i_0<\ldots<i_n\le r} n_{i_0}\cdots n_{i_n} V^2(x_{i_0},\ldots,x_{i_n}).$$

This relationship immediately reveals that ξ_0^N must include ± 1 in its support. For convenience, set $x_0^N = -x_n^N = -1$.

Now application of Theorem 2.1 implies that $r \le 2n$ for a D-optimal exact design ξ_0^N and that it can have no more than n-1 interior clusters. Thus it may be assumed that $n_i^N = 1$ for $i=n+1,\ldots,r$. For convenience it may also be assumed that $x_1^N < \ldots < x_{n-1}^N$.

The following lemma is essentially a statement that ξ_0^N converges weakly to ξ_0.

<u>Lemma 3.1</u>: $x_i^N \to x_i^0$ and $n_i^N/N \to 1/(n+1)$ as $n \to \infty$ for $i=0,\ldots,n$.

<u>Proof:</u> First note that $\xi_*^N \xrightarrow{w} \xi_0$. Also, $|M(\xi_0)| \ge |M(\xi_0^N)| \ge |M(\xi_*^N)| \to |M(\xi_0)|$. Therefore, $|M(\xi_0^N)| \to |M(\xi_0)|$. The proof will be complete once it is established that $\xi_0^N \xrightarrow{w} \xi_0$.

Suppose that ξ_0^N did not converge weakly to ξ_0. Then there exists a continuity point y_0 of F_0 such that $F_N(y_0)$ does not converge to $F_0(y_0)$, where $F_N(y) = \xi_0^N([-1,y])$ and $F_0(y) = \xi_0([-1,y])$. Thus there exists $\varepsilon>0$ and a sequence $\{N_j\}$ such that $|F_{N_j}(y_0) - F_0(y_0)| > \varepsilon$ for all j. According to the Helly selection theorem, there exists a subsequence $\{N_{j_k}\}$ and a measure $\tilde{\xi}$ such that $\xi_0^{(N_{j_k})} \xrightarrow{w} \tilde{\xi}$. Therefore $M(\xi_0^{(N_{j_k})}) \longrightarrow M(\tilde{\xi})$ so that $|M(\xi_0^{(N_{j_k})})| \longrightarrow |M(\tilde{\xi})|$. Since $|M(\xi_0^N)| \longrightarrow |M(\xi_0)|$ has already been established and since ξ_0 is unique, this implies that $\tilde{\xi} = \xi_0$. Hence $F_{N_{j_k}}(y_0) \longrightarrow F_0(y_0)$. This contradiction implies that $\xi_0^N \xrightarrow{w} \xi_0$ and completes the proof.

The following lemma will be needed for the proof of Theorem 3.1.

Lemma 3.2: If $i \neq \ell$, then

$$\frac{\partial}{\partial x_i}\left\{\psi_N\begin{bmatrix}x_0,\ldots,x_k\\ n_0,\ldots,n_k\end{bmatrix}\right\}_{x_i=x_\ell} = n_i\frac{\partial}{\partial x_\ell}\left\{\psi_{N-n_i-n_\ell+1}\begin{bmatrix}x_j\\ n_j,\ j\neq i;n_\ell=1\end{bmatrix}\right\}. \qquad (3.1)$$

Proof: According to the Binet-Cauchy formula, the left hand side of (3.1) is

$$\frac{\partial}{\partial x_i}\left\{\sum_{1\leq i_0<\ldots<i_n\leq k} n_{i_0}\cdots n_{i_n}\, V^2(x_{i_0},\ldots x_{i_n})\right\}_{x_i=x_\ell}. \qquad (3.2)$$

First note that the summation need only be taken over sequences $i_0<\ldots<i_n$ which include i. If the sequence also includes ℓ, then $V^2(x_{i_0},\ldots,x_{i_n})$, as a polynomial in x_i, has a double root at $x_i = x_\ell$. Therefore such sequences may be deleted from the summation and (3.2) equals

$$\frac{\partial}{\partial x_i}\left\{\sum_{\substack{1\leq i_1<\ldots<i_n\leq k\\ i_s\neq i,\ell \text{ for } s=1,\ldots,n}} n_i n_{i_1}\cdots n_{i_n}\, V^2(x_i,x_{i_1},\ldots,x_{i_n})\right\}_{x_i=x_\ell}$$

$$= n_i\frac{\partial}{\partial x_\ell}\left\{\sum_{\substack{1\leq i_1<\ldots<i_n\leq k\\ i_s\neq i \text{ for } s=1,\ldots,n}} n_{i_1}\cdots n_{i_n}\, V^2(x_\ell,x_{i_1},\ldots,x_{i_n})\right\}$$

$$= n_i\frac{\partial}{\partial x_\ell}\left\{\psi_{N-n_i-n_\ell+1}\begin{bmatrix}x_j\\ n_j,j\neq i;n_{\ell=1}\end{bmatrix}\right\}.$$

These preliminary results will now be used to prove the following theorem.

Theorem 3.1: For N sufficiently large, a D-optimal exact design $\xi_0^N = \xi_*^N$.

Proof: Suppose that the conclusion of the theorem were false. Then there would exist an integer k such that $n<k\leq 2n$ and such that ξ_0^N has k+1 support points for infinitely many values of N. Several steps are needed to show that this supposition cannot hold.

i. Recall first from Lemma 3.1 that $x_i^N \longrightarrow x_i^0$ and $n_i^N/N \longrightarrow 1/(n+1)$ for $i=0,\ldots,n$.

For the remainder of the proof, N will be assumed to be one of the infinitely many values for which ξ_0^N has k+1 support points.

ii. The limiting behavior of x_i^N will now be considered for $n<i\leq k$. This sequence is bounded by $\pm$ 1 and so a limit point x_i^0 exists. It will be shown that $x_i^0 = x_j^0$ for some $j\in\{1,\ldots,n-1\}$.

Since ξ_0^N maximizes $|M(\xi^N)|$, it must be true that

$$0 = \frac{1}{N^n} \frac{\partial\psi_N}{\partial x_i} \begin{bmatrix} x_0^N,\ldots,x_k^N \\ n_0^N,\ldots,n_K^N \end{bmatrix}$$

$$= \sum_{1\leq i_0<\ldots<i_n\leq k} \frac{n_{i_0}^N\ldots n_{i_n}^0}{N^n} \frac{\partial V^2}{\partial x_i} (x_{i_0}^N,\ldots,x_{i_n}^N)$$

$$= \sum_{1\leq i_1<\ldots<i_n\leq k} \frac{n_{i_1}^N\ldots n_{i_n}^N}{N^n} \frac{\partial V^2}{\partial x_i} (x_{i_1}^N,\ldots,x_{i_n}^N,x_i^N). \tag{3.3}$$

Recall now that $n_i=1$ for $n<i\leq k$. Thus the summands in (3.3) vanish in the limit as $N\to\infty$ unless $i_n \leq n$. Therefore, taking the limit of (3.3) and applying Lemma 3.1 yields

$$0 = \sum_{1\le i_1<\dots<i_n\le n} \left(\frac{1}{n+1}\right)^n \frac{\partial V^2}{\partial x_i}(x^0_{i_1},\dots,x^0_{i_n},x^0_i)$$

$$= \left(\frac{1}{n+1}\right)^n \frac{\partial\psi_{n+2}}{\partial x_i}\begin{bmatrix} x^0_0,\dots,x^0_n,x^0_i \\ 1,\dots,1,1\end{bmatrix} . \tag{3.4}$$

Applying exactly the same methods to the condition

$$\frac{1}{N^n}\frac{\partial^2\psi_N}{\partial X_i^2}\begin{bmatrix} x^N_0,\dots,x^N_k \\ n^N_0,\dots,n^N_k\end{bmatrix} \le 0$$

yields

$$\left(\frac{1}{n+1}\right)^n \frac{\partial^2\psi_{n+2}}{\partial x_i^2}\begin{bmatrix} x^0_0,\dots x^0_n,x^0_i \\ 1,\dots,1,1\end{bmatrix} \le 0 . \tag{3.5}$$

As a function of x_i, $\psi_{n+2}\begin{bmatrix} x^0_0,\dots,x^0_n,x_i \\ 1,\dots,1,1\end{bmatrix}$

is a polynomial of degree 2n. For $0\le\ell\le n$, the Binet-Cauchy formula implies that $\psi_{n+2}\begin{bmatrix} x^0_0,\dots,x^0_n,x^0_\ell \\ 1,\dots,1,1\end{bmatrix} = 2V^2(x^0_0,\dots x^0_n)$. That is, this polynomial assumes the same value at the n+1 points $-1=x^0_0<x^0_1<\dots<x^0_{n-1}<x^0_n=1$. Therefore $\frac{\partial\psi_{n+2}}{\partial x_i}$ must have at least one root in each of the n intervals of the form $(x^0_\ell,x^0_{\ell+1})$, where $0\le\ell\le n-1$.

Furthermore, application of Lemma 3.2 yields

$$\frac{\partial\psi_{n+2}}{\partial x_i}\begin{bmatrix} x^0_0,\dots,x^0_n,x^0_\ell \\ 1,\dots,1,1\end{bmatrix} = \frac{\partial V^2}{\partial x_\ell}(x^0_0,\dots,x^0_n) = 0$$

for $\ell=1,\dots,n-1$. Here the equality to zero follows from the (approximate theory) result that $-1=x^0_0<x^0_1<\dots<x^0_{n-1}<x^0_n=1$ maximize V^2.

Now note that, according to the Binet-Cauchy formula, decreasing (increasing) $x_0^0(x_n^0)$ would increase ψ_{n+2}. Therefore $\frac{\partial\psi_{n+2}}{\partial x_i}$ must be negative (positive) at $x_0^0(x_n^0)$.

The net result of these properties is that ψ_{n+2} has one root in each interval of the form $(x_\ell^0, x_{\ell+1}^0)$, where $0 \leq \ell \leq n-1$, and at that root $\frac{\partial^2\psi_{n+2}}{\partial x_i^2} > 0$. Therefore (3.4) and (3.5) imply that x_i^0 must be one of the points $x_1^0, \ldots, x_{n-1}^0$.

iii. The main idea of the proof is to exploit the following Taylor series expansion.

$$\psi_N \begin{bmatrix} x_0^N, \ldots, x_k^N \\ n_0^N, \ldots, n_k^N \end{bmatrix} = \psi_N \begin{bmatrix} -1, & x_1^N, \ldots x_{n-1}^N, & 1, & x_{n+1}^N, \ldots, x_k^N \\ n_0^N, & n_1^N, \ldots n_{n-1}^N, & n_n^N, & 1 \quad, \ldots, 1 \end{bmatrix}$$

$$= \psi_N \begin{bmatrix} x_0^0, \ldots, x_k^0 \\ n_0^N, \ldots, n_k^N \end{bmatrix}$$

$$+ \sum_{i \neq 0,n} \frac{\partial\psi_N}{\partial x_i} \begin{bmatrix} x_0^0, \ldots, x_k^0 \\ n_0^N, \ldots, n_k^N \end{bmatrix} (x_i^N - x_i^0)$$

$$+ \frac{1}{2} \sum_{i,j \neq 0,n} \frac{\partial^2\psi_N}{\partial x_i \partial x_j} \begin{bmatrix} \tilde{x}_0^N, \ldots, \tilde{x}_k^N \\ n_0^N, \ldots, n_k^N \end{bmatrix} (x_i^N - x_i^0)(x_j^N - x_j^0), \qquad (3.6)$$

where each $\tilde{x}_i^N$ lies between x_i^N and x_i^0. It will subsequently be shown that the first order terms vanish and that the second order term is negative for N sufficiently large. Once demonstrated, these results will imply that $\psi_N \begin{bmatrix} x_0^0, \ldots, x_k^0 \\ n_0^N, \ldots, n_k^N \end{bmatrix} > \psi_N \begin{bmatrix} x_0^0, \ldots, x_k^N \\ n_0^N, \ldots, n_k^N \end{bmatrix}$

for N sufficiently large. This contradiction of the D-optimality of ξ_0^N must imply that k=n for N sufficiently large which will complete the proof of the theorem.

iv. To show that the first order terms in (3.6) vanish, consider the case that $1 \le i \le n-1$ and that $x_j^0 \ne x_i^0$ for $j=n+1,\ldots,k$. For $0 \le \ell \le n$, let p_ℓ denote n_ℓ^N plus the number of points among $x_{n+1}^0,\ldots,x_k^0$ which equal x_ℓ^0. Then

$$\frac{\partial \psi_N}{\partial x_i}\begin{bmatrix} x_0^0,\ldots,x_k^0 \\ n_0^N,\ldots,n_k^N \end{bmatrix} = \frac{\partial}{\partial x_i}\left\{ p_0 \cdots p_n \; V^2(x_0^0,\ldots,x_{i-1}^0,x_i,x_{i+1}^0,\ldots,x_n^0)\right\}_{x_i=x_i^0}$$

$$= 0$$

since $-1 = x_0^0 < x_1^0 < \ldots < x_{n-1}^0 < x_n^0 = 1$ maximize V^2.

Consider next the case that $1 \le i \le n-1$ and that exactly r of the points $x_{n+1}^0,\ldots,x_k^0$ equal x_i^0. $(r=p_i-n_i^N.)$ Application of Lemma 3.2 yields

$$\frac{\partial \psi_N}{\partial x_i}\begin{bmatrix} x_0^0,\ldots,x_k^0 \\ n_0^N,\ldots,n_k^N \end{bmatrix} = \frac{\partial}{\partial x_i}\left\{ \psi_N \begin{bmatrix} x_0^0,\ldots,x_{i-1}^0,x_i,x_{i+1}^0,\ldots,x_n^0,x_i^0 \\ p_0,\ldots,p_{i-1},n_i^N,p_{i+1},\ldots,p_n,r \end{bmatrix}\right\}_{x_i=x_i^0}$$

$$= n_i^N \frac{\partial}{\partial y}\left\{ \psi_{N-p_i+1} \begin{bmatrix} x_0^0,\ldots,x_{i-1}^0,y,x_{i+1}^0,\ldots,x_n^0 \\ p_0,\ldots,p_{i-1},1,p_{i+1},\ldots,p_n \end{bmatrix}\right\}_{y=x_i^0}$$

$$= \frac{n_i^N}{p_i}\; p_0 \cdots p_n \frac{\partial}{\partial y}\left\{ V^2\,(x_0^0,\ldots,x_{i-1}^0,y,x_{i+1}^0,\ldots,x_n^0)\right\}_{y=x_i^0}$$

$$= 0.$$

The final case is that $n+1 \le i \le k$. In this case, there exists $j \in \{1,\ldots,n-1\}$ such that $x_i^0 = x_j^0$.

Hence Lemma 3.2 again yields

$$\frac{\partial\psi_N}{\partial x_i}\begin{bmatrix} x_0^0,\ldots,x_k^0 \\ n_0^N,\ldots,n_k^N \end{bmatrix} = \frac{\partial}{\partial x_i}\left\{\psi_N\begin{bmatrix} x_0^0,\ldots,x_{j-1}^0,x_j^0,x_{j+1}^0,\ldots,x_n^0,x_i \\ p_0,\ldots,p_{j-1},p_j-1,p_{j+1},\ldots,p_n,1 \end{bmatrix}\right\}_{x_i=x_j^0}$$

$$= \frac{\partial}{\partial y}\left\{\psi_{N-p_j+1}\begin{bmatrix} x_0^0,\ldots,x_{j-1}^0,y,x_{j+1}^0,\ldots,x_n^0 \\ p_0,\ldots,p_{j-1},1,p_{j+1},\ldots,p_n \end{bmatrix}\right\}_{y=x_j^0}$$

$$= \frac{p_0\cdots p_n}{p_j}\frac{\partial}{\partial y}\left\{V^2(x_0^0,\ldots,x_{j-1}^0,y,x_{j+1}^0,\ldots,x_n^0)\right\}_{y=x_j^0}$$

$$= 0 \, .$$

The net result is that $\frac{\partial\psi_N}{\partial x_i}\begin{bmatrix} x_0^0,\ldots,x_k^0 \\ n_0^N,\ldots,n_k^N \end{bmatrix} = 0$

for $i\neq 0,n$. Thus the first order terms in (3.6) all vanish. Note that this result holds for any value of N.

v. The final step of the proof is to demonstrate that the matrix Ψ^N is negative definite for N sufficiently large, where the elements of Ψ^N are $\frac{\partial^2\psi_N}{\partial x_i\partial x_j}\begin{bmatrix} \tilde{x}_0^N,\ldots,\tilde{x}_k^N \\ n_0^N,\ldots,n_k^N \end{bmatrix}$.

Here i, $j=1,\ldots,n-1,n+1,\ldots,k$. It will be convenient to use the notation

$$\Psi^N = \begin{bmatrix} A^N & B^N \\ B^{N'} & C^N \end{bmatrix},$$

where A^N is $(n-1)\times(n-1)$ and C^N is $(k-n)\times(k-n)$. The negativity of Ψ^N will be established by showing that its i^{th} principal minor D_i^N satisfies $(-1)^i D_i^N > 0$ for N sufficiently large and for $i=1,\ldots,n-1,n+1,\ldots,k$. (This particular set of k-1 indices is chosen to correspond to previous notation.)

It will first be shown that $(-1)D_i^N > 0$ for $i=1,\ldots,n-1$ and for N sufficiently large. To obtain this result, it suffices to show that A^N is negative definite for N sufficiently large. Equivalently, it will be shown that A^N/N^{n+1} is negative definite for large N. Note first that

$$\left(\frac{A^N}{N^{n+1}}\right)_{ij} = \sum_{1\leq i_0<\ldots<i_n\leq k} \frac{n^N_{i_0}\ldots n^N_{i_n}}{N^{n+1}} \frac{\partial^2}{\partial x_i \partial x_j}\left\{V^2(x_{i_0},\ldots,x_{i_n})\right\}_{x_\ell=\tilde{x}^N_\ell;\ell=0,\ldots,k}$$

$$\to \left(\frac{1}{n+1}\right)^{n+1} \frac{\partial^2}{\partial x_i \partial x_j}\left\{V^2(x_0,\ldots,x_n)\right\}_{x_\ell = x^0_\ell;\ell=0,\ldots,n.}$$

The convergence follows because $n^N_i=1$ for $i=n+1,\ldots,k$. Let A^0 denote the limiting matrix thus obtained. Then A^0 is negative definite due to the (approximate theory) result that $-1=x^0_0<x^0_1<\ldots<x^0_{n-1}<x^0_n = 1$ maximize V^2.

It may now be shown that the negativity of A^0 implies that A^N/N^{n+1} is negative definite for N sufficiently large. First let $\lambda^0_{n-1}<0$ denote the largest eigenvalue of A^0 and let $\Delta^N = A^N/N^{n+1} - A^0$ for each N. It has been established that $\Delta^N\to 0$ and so there exists N_0 such that $\max_{1\leq i,j\leq n-1} |\Delta^N_{ij}|\leq -\frac{1}{2}\lambda^0_{n-1}/(n-1)^2$ whenever $N\geq N_0$. Then for any $y\in\mathbb{R}^{n-1}$, $N\geq N_0$ implies that

$$\begin{aligned} y'\Delta^N y &= \sum_{i,j=1}^{n-1} \Delta^N_{ij}\, y_i y_j \\ &\leq (n-1)^2 \max_{1\leq i,j\leq n-1} |\Delta^N_{ij}| \max_{1\leq i\leq n-1} |y_i|^2 \\ &\leq -\frac{1}{2}\lambda^0_{n-1}\, y'y. \end{aligned}$$

Therefore,

$$\begin{aligned} y' \frac{A^N}{N^{n+1}} y &= y'A^0 y + y'\Delta^N y \\ &\leq (\lambda^0_{n-1} - \frac{1}{2}\lambda^0_{n-1})\, y'y = \frac{1}{2}\lambda^0_{n-1}\, y'y \qquad (3.7) \end{aligned}$$

whenever $N \geq N_0$. The right hand side of (3.7) is negative unless y=0. Therefore A^N/N^{n+1} is negative definite for $N \geq N_0$.

It now remains only to show that $(-1)^i D_i^N > 0$ for i=n+1,...,k and for N sufficiently large. For purposes of notation, let $D_i^N = \begin{vmatrix} A^N & \bar{B}^N \\ \bar{B}^{N'} & \bar{C}^N \end{vmatrix}$,

where $\bar{C}^N$ is (i-n)x(i-n). It has already been established that A^N is negative definite for large N. Hence $D_i^N = |A^N| \; |\bar{C}^N - \bar{B}^{N'}(A^N)^{-1}\bar{B}^N|$ and it suffices to show that the matrix $\bar{F}^N = \{\bar{C}^N - \bar{B}^{N'}(A^N)^{-1}\bar{B}^N\}/N^n$ is negative definite for sufficiently large N. For $n+1 \leq j,m \leq i$,

$$\left(\frac{\bar{F}^N}{N^n}\right)_{j,m} = \sum_{1 \leq i_0 < \ldots < i_n \leq k} \frac{n_{i_0}^N \ldots n_{i_n}^N}{N^n} \frac{\partial^2}{\partial x_j \partial x_m} \left\{ V^2(x_{i_0}, \ldots, x_{i_n}) \right\}_{x_\ell = \tilde{x}_\ell^N;\ \ell=0,\ldots,k.} \tag{3.8}$$

Here $n_j^N = n_m^N = 1$ and so the limit of (3.8) is zero unless j=m. For the diagonal elements,

$$\left(\frac{\bar{F}^N}{N^n}\right)_{j,j} \to \left(\frac{1}{n+1}\right)^{n+1} \sum_{p=0}^{n} \frac{\partial^2}{\partial x_j^2} \left\{ V^2(x_0^0, \ldots, x_{p-1}^0, x_{p+1}^0, \ldots, x_n^0, x_j) \right\}_{x_j = x_j^0}. \tag{3.9}$$

If $L_0(x),\ldots,L_n(x)$ denote the Lagrange polynomials such that $L_i(x_j^0) = \delta_{ij}$ and the vector $\mathscr{L}_p(x_j) = [L_0(x_j),\ldots,L_{p-1}(x_j), L_{p+1}(x_j),\ldots,L_n(x_j)]'$, then

$$V^2(x_0^0,\ldots,x_{p-1}^0,x_{p+1}^0,\ldots,x_n^0,x_j) = \prod_{0 \leq r < s \leq n} (x_s^0 - x_r^0)^2 \begin{vmatrix} I_n & \mathscr{L}_p(x_j) \\ 0 & L_p(x_j) \end{vmatrix}^2$$

$$= L_p^2(x_j) \prod_{0 \leq r < s \leq n} (x_s^0 - x_r^0)^2.$$

Therefore, the right hand side of (3.9) becomes

$$\left(\frac{1}{n+1}\right)^{n+1} \prod_{0\leq r<s\leq n} (x_s^0 - x_r^0)^2 \frac{\partial^2}{\partial x_j^2} \left\{ \sum_{p=0}^{n} L_p^2(x_j) \right\}_{x_j=x_j^0}. \tag{3.10}$$

Recall now that (according to the approximate design result on D and G-optimality for polynomial regression) $d(x,\xi_0)=(n+1)\sum_{p=0}^{n} L_p^2(x)$ and $\frac{\partial^2}{\partial x^2}\left\{d(x,\xi_0)\right\}_{x=x_\ell^0} < 0$ for $\ell=1,\ldots,n-1$.

Therefore the expression in (3.10) is strictly negative so that $\bar{F}^N/n^n \to F^0$, a negative definite diagonal matrix.

Applying the same argument to $\bar{F}^N/N^n$ which was applied to A^N/N^{n+1} yields the conclusion that there exists N_i such that $\bar{F}^N/N^n$ is negative definite whenever $N\geq N_i$.

The net result is that whenever $N\geq\max[N_0,N_{n+1},\ldots,N_k]$, then Ψ^N is negative definite. This completes the final step of the proof of the theorem.

In the case of linear regression, the validity of Conjecture 3.1 is readily demonstrated. Of course N=2k implies that $\xi_0^N = (\delta_{-1}+\delta_1)/2 = \xi_*^N$. For N=2k+1, Theorem 2.1 implies that only exact designs of the form $\xi^N = [(k-s)\delta_{-1} + \delta_x + (k+s)\ \delta_1]/N$, where $0\leq s\leq k$ and $-1\leq x\leq 1$, need be considered. Now it is not hard to show (by elementary calculations) that the exact designs $\xi_1^N = [k\delta_{-1}+(k+1)\delta_1]/N$ and $\xi_2^N = [(k+1)\delta_{-1}+k\delta_1]/N$ are both D-optimal and satisfy Conjecture 3.1.

Federov (1972) suggests that the conjecture holds for n=3.

For n=2, numerical work has been done to determine the D-optimal exact design of the form $\xi^N = (n_0\delta_{-1} + \delta_{y_1} + n_2\delta_{y_2} + \delta_{y_3} + n_4\delta_1)/N$, where $-1 \leq y_1 \leq y_2 \leq y_3 \leq 1$ and $n_0 + 1 + n_2 + 1 + n_4 = N$.

For each possible choice of the integers n_0, n_2, n_4, the PUCC subroutine SECANT was utilized in a Fortran program to solve the system of non-linear equations obtained by setting the partial derivatives of $|M(\xi^N)|$ with respect to y_1, y_2, and y_3 equal to 0. The nature of the procedure requires that $N \geq 7$. The following table displays the best designs thus obtained.

Table 3.1: Exact Designs for Quadratic Regression on [-1,1]

N	$\tilde{\xi}^N$	$\lvert M(\xi^N)\rvert$
7	$\begin{bmatrix} -1, & .0000, & +1 \\ 2, & 3, & 2 \end{bmatrix}$	.1399
8	$\begin{bmatrix} -1, & .000, & +1 \\ 3, & 3, & 2 \end{bmatrix}$	.1406
	$\begin{bmatrix} -1, & .0000, & +1 \\ 2, & 3, & 3 \end{bmatrix}$	.1406
9	$\begin{bmatrix} -1, & .0, & +1 \\ 3, & 3, & 3 \end{bmatrix}$	.1481
10	$\begin{bmatrix} -1, & .000, & +1 \\ 4, & 3, & 3 \end{bmatrix}$	.1440
	$\begin{bmatrix} -1, & .00, & +1 \\ 3, & 4, & 3 \end{bmatrix}$	.1440
	$\begin{bmatrix} -1, & .000, & +1 \\ 3, & 3, & 4 \end{bmatrix}$	.1440
11	$\begin{bmatrix} -1, & .0000, & +1 \\ 4, & 4, & 3 \end{bmatrix}$	.1443
	$\begin{bmatrix} -1, & .000, & +1 \\ 4, & 3, & 4 \end{bmatrix}$	.1443
	$\begin{bmatrix} -1, & .00, & +1 \\ 3, & 4, & 4 \end{bmatrix}$	.1443

Note that these calculations invariably gave support $(\tilde{\xi}^N) = \{-1,0,+1\}$ and $\tilde{\xi}^N = \tilde{\xi}^N_*$.

Similar work has been done for n=3. The following table displays the best exact designs thus obtained. These should be compared to ξ^N_* which distributes observations as evenly as possible among the points $\pm$ 1, $\pm$ $1/\sqrt{5}$.

Table 3.2: Exact Designs for Cubic Regression on [-1,1]

N	$\tilde{\xi}^N$	$\|M(\tilde{\xi}^N)\|$
4	$\begin{bmatrix} -1, & -.447213596, & .447213596, & +1 \\ 1, & 1 \quad , & 1 \quad , & 1 \end{bmatrix}$	.0051200
5	$\begin{bmatrix} -1, & -.447213596, & .447213596, & +1 \\ 1, & 2 \quad , & 1 \quad , & 1 \end{bmatrix}$	.0041943
	$\begin{bmatrix} -1, & -.447213596, & .447213596, & +1 \\ 1, & 1 \quad , & 2 \quad , & 1 \end{bmatrix}$	.0041943
6	$\begin{bmatrix} -1, & -.44721360, & .44721360, & +1 \\ 1, & 2 & 2 \quad , & 1 \end{bmatrix}$	.0040454

Here it is seen that $\tilde{\xi}^N = \xi^N_*$ in each case. Again, the validity of Conjecture 3.1 is suggested.

4. G-Optimality

As already noted, a G-optimal exact design ξ^N_0 satisfies

$$\max_{\mathscr{X}} d(x,\xi^N_0) = \min_{\xi^N \in \Xi_N} \max_{\mathscr{X}} d(x,\xi^N),$$ where the variance function

$$d(x,\xi^N) = f(x)'M^{-1}(\xi^N)f(x).$$

Guest (1958) obtained the G-optimal approximate design ξ_0 for polynomial regression of degree on [a,b] = [-1,1]. This later turned out to coincide with D-optimal approximate design given by Hoel (1958), leading Kiefer and Wolfowitz (1960) to prove that the two criteria are equivalent in the general approximate design setting.

For polynomial regression of degree n, the G-optimal exact design coincides with the G(and D)-optimal approximate design when N is a multiple of n+1. Otherwise, G-optimal exact designs can exhibit some interesting behavior as may be seen in the following examples.

Example 4.1: Consider the most simple example of linear regression on [-1,1]. Here $d(x,\xi^N) = 1 + (x-\mu_1)^2/(\mu_2-\mu_1^2)$ and so

$$\max_{-1\le x\le 1} d(x,\xi^N) = \begin{cases} 1 + (\mu_1-1)^2/(\mu_2-\mu_1^2) & -1\le\mu_1\le 0 \\ 1 + (\mu_1+1)^2/(\mu_2-\mu_1^2) & 0\le\mu_1\le 1. \end{cases}$$

Therefore, a G-optimal exact design ξ_0^N will have $\mu_1=0$ and will maximize μ_2 among exact designs satisfying $\mu_1=0$. Thus

$$\xi_0^N = \begin{cases} (\delta_{-1} + \delta_1)/2 & N=2k \\ (k\delta_{-1} + \delta_0 + k\delta_1)/2 & N=2k+1. \end{cases}$$

Note that

$$\max_{-1\le x\le 1} d(x,\xi_0^N) = \begin{cases} 2 & N=2k \\ 2 + 1/(N-1) & N=2k+1 \end{cases}$$

whereas $\max_{-1\le x\le 1} d(x,\xi_0) = 2$. Note also that for N=2k+1, the design ξ_0^N always has an interior singlet.

Example 4.2: Consider the setting of quadratic regression on [-1,1]. For N=3k, the G-optimal exact design is $\xi_0^N = \xi_0 = (\delta_{-1} + \delta_0 + \delta_1)/3$.

For N=3k+1, it is believed that the form of the G-optimal exact design is $\xi_0^N = [k\delta_{-1} + \delta_{-u} + (k-1)\delta_0 + \delta_u + k\delta_1]/N$. Among such designs, G-optimality will be attained if and only if $d(0,\xi_0^N) = d(1,\xi_0^N)$. Manipulation of this condition yields

$$(3-5/k)u^4 - (9-1/k)u^2 + 2 = 0.$$

An interesting consequence of this result is that $u^2 \to (9-\sqrt{57})/6$ as $k \to \infty$. That is, for large k, the G-optimal exact design for N=3k+1 has singlets at approximately $\pm$ .4916. Perhaps even more interesting is that $u^2 = \sqrt{5} - 2$ for k=1. Thus , for k=1, the two singlets $\pm u \approx \pm .4859$ are already very close to their asymptotic values. Note that

$$\max_{-1\leq x\leq 1} d(x,\xi_0^N) = 1 + \{-1+(3k+1)(k+u^4)/2(k+u^2)^2\}^{-1} \quad \text{whereas}$$

$$\max_{-1\leq x\leq 1} d(x,\xi_0) = 3.$$

For N=3k+2, it is believed that the form of the G-optimal exact design is $\xi_0^N = (k\delta_{-1} + \delta_{-v} + k\delta_0 + \delta_v + k\delta_1)/N$. Among such designs, ξ_0^N will be G-optimal if and only if $d(0,\xi_0^N) = d(1,\xi_0^N)$. Therefore,

$$(3-4/k)v^4-(9-2/k)v^2 + 4 = 0.$$

In the limit, $v^2 \to (9-\sqrt{33})/6$ and the singlets converge to $\pm V \approx \pm .7366$. For k=1, the singlets are at $\pm v = \pm .7288$ which are already close to the asymptotic values. Note finally that

$$\max_{-1\leq x\leq 1} d(x,\xi_0^N) = 1 + \{-1+(3k+2)(k+v^4)/2(k+v^2)^2\}^{-1}.$$

The one gap in this example would be filled by the proof of the following conjecture.

<u>Conjecture 4.1</u>: A G-optimal exact design for polynomial regression on [-1,1] must be symmetrical about the origin.

For $n \geq 3$ and $N \neq 0 \bmod(n+1)$, it may be seen that the clusters of ξ_0^N will not coincide with the support points of ξ_0.

References

Federov, V. V. (1972). Theory of Optimal Experiments. Translated by Studden, W.J. and Klimko, E. M., Academic Press, New York.

de la Garza, A. (1954). Spacing of information in polynomial regression. Ann. Math. Statist., Vol. 25, 123-130.

Guest, P.G. (1958). The spacing of observations in polynomial regression. Ann. Math. Statist., Vol. 29, 294-299.

Hoel, P.G. (1958). Efficiency problems in polynomial estimation. Ann. Math. Statist., Vol. 29, 1134-1145.

Karlin, S. and Studden, W.J. (1966). Tchebycheff Systems: with Applications in Analysis and Statistics. Interscience Publishers, New York.

Kiefer, J. (1959). Optimum experimental designs. J. Roy. Statist. Soc. Ser. B, Vol. 21, 273-319.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence to two extremum problems. Can. J. Math., Vol. 12, 363-366.

Salaevskii, O.V. (1966). The problem of the distribution of observation in polynomial regression. Proc. Steklov Inst. Math. #79, 147-166.