THEORY OF CANONICAL MOMENTS AND ITS APPLICATIONS IN POLYNOMIAL REGRESSION - PART II*

by

Tai Shing Lau
Purdue University

Technical Report #83-24

Department of Statistics
Purdue University

August 1983

*This research was supported by NSF Grant No. MCS-8200631 and MCS 81-0167-A1

CHAPTER III

D-OPTIMAL DESIGN

3.1 General Results

Consider the model

$$y(x) = \sqrt{w(x)}\,(\beta_0 + \beta_1 x + \ldots + \beta_m x^m) + \varepsilon$$
$$= \beta^T f(x) + \varepsilon \qquad (3.1.1)$$

where $x \in [a,b]$, $f^T(x) = \sqrt{w(x)}\,(1,x,\ldots,x^m)$ and $E(\varepsilon) = 0$, $\mathrm{var}(\varepsilon) = \sigma^2$.

The information matrix is given by

$$M(\xi) = \int_a^b f(x)f^T(x)d\xi.$$

A design ξ^* is said to be weighted D-optimal if ξ^* maximizes $|M(\xi)|$, the determinant of the information matrix.

Remark 3.1.1: It should be noted that (3.1.1) is a generalization of the model introduced in (1.1.1) in that it is assumed that $\mathrm{var}(\varepsilon) = \frac{\sigma^2}{w(x)}$ in (1.1.1), where $w(x)$ is a known nonnegative continuous function. That is, the assumption of homogeneity of variance, which is an important postulate in the ordinary least squares method, is violated. We can modify the model by the following transformations:

$$f(x) = \sqrt{w(x)}\,(1,x,\ldots,x^m),$$
$$\tilde{y} = \sqrt{w(x)}\,y.$$

It can be easily checked that $\mathrm{var}(\tilde{y}) = \sigma^2$. So we have the variance homogenized and the resulting model is exactly (3.1.1).

The D-optimal design will stay unchanged under a linear transformation on $f(x)$.

Theorem 3.1.1: (Kiefer). Let T be a non-singular matrix of size m+1. The D-optimal design for $f(x)$ and the D-optimal design for $Tf(x)$ are the same.

Proof: See Kiefer (1959).

Corollary 3.1.1. Let $x^* = cx+d$, $c > 0$ be a mapping from $[a,b]$ to $[e,f]$. The D-optimal design on $[e,f]$ is the same as the D-optimal design on x, assuming $w(x)$ undergoes the same transformation.

Proof: See Kiefer (1959).

The immediate result of Corollary 3.1.1 is that the D-optimal design remains unchanged if we use different units in measuring x.

Next we will show that the D-optimal design has another desirable property. Let us introduce the following definitions.

Definition 3.1.1: The variance function of a design ξ at a point x is given by

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

provided that $M^{-1}(\xi)$ exists. (Notice that the "true" variance of the least squares estimator of $\beta^T f(x)$ at a point x is given by $\frac{\sigma^2}{N} d(x,\xi)$ where N is the sample size.)

Definition 3.1.2: The design that minimizes $\max_x d(x,\xi)$ is called the weighted G-optimal design.

Suppose we define $g(x) = T f(x)$, where T is a nonsingular matrix of size m+1. It is easy to check that the variance function of ξ associated with f(x) remains the same if we replace f(x) by g(x). Thus we have proved

Theorem 3.1.2: Let T be a nonsingular matrix of size m+1. The G-optimal design for f(x) coincides with the G-optimal design for T f(x).

Immediately, we have the following corollary.

Corollary 3.1.2: Let $x^* = cx+d$, $c > 0$, be a mapping from [a,b] to [e,f]. The G-optimal design on [e,f] is the same as the G-optimal design on [a,b].

Corollary 3.1.2 implies that G-optimal design remains unchanged if we use different units in measuring x.

Guest (1958) and Hoel (1958) found the D-optimal design and G-optimal design respectively and their results show that the D-optimal design coincides with the G-optimal design. The coincidence led Kiefer and Wolfowitz (1960) to prove their celebrated equivalence theorem. The following form of the equivalence theorem can be found in Federov (1972).

Theorem 3.1.3: The following are equivalent:

1. The design ξ^* maximizes $|M(\xi)|$.
2. The design ξ^* minimizes $\max_x d(x,\xi)$.

3. $\max_x d(x,\xi^*) = m+1$.

It follows that the weighted G-optimal designs are equivalent to weighted D-optimal designs. The D-optimal design looks more appealing now since it can be interpreted as the design that minimizes the maximum of the variance function. It is no wonder that the D-optimal designs are used extensively. The D-optimal criterion has been applied in different settings: polynomial regression on the n-dim cube, the simplex and the sphere, polynomial regression on the interval with classical weight functions; trigonometric regression on the circle, etc. For more details see Kiefer (1959, 1961a, 1961b), Kiefer and Walbran (1967), Karlin and Studden (1966a). Federov (1972) initiated the study of algorithms for finding the optimal designs. His approach gives a practical way to approximate the D-optimal design in a more general setting. For other approaches to D-optimal design see Silvey and Titterington (1974) and Pukelsheim (1980).

3.2 Some Admissibility Results

A design ξ is called admissible if there does not exist a design ξ' such that $M(\xi') \geq M(\xi)$. The inequality means that $M(\xi') - M(\xi)$ is positive semi-definite and $M(\xi') \neq M(\xi)$. The proof of the following lemma can be found in Karlin and Studden (1966a).

Lemma 3.2.1: Let $w(x) \equiv 1$. $M(\xi') \geq M(\xi)$ and $M(\xi') \neq M(\xi)$ iff $\mu_i(\xi') = \mu_i(\xi)$ for $0 \leq i \leq 2m-1$ and $\mu_{2m}(\xi') > \mu_{2m}(\xi)$. Here $\mu_i(\alpha)$ denotes the i-th moment of the measure α, i.e. $\mu_i(\alpha) = \int_a^b x^i d\alpha$.

Immediately, we have the following theorem

Theorem 3.2.1: The design ξ is admissible for polynomial regression of degree m ($w(x) \equiv 1$) iff $p_{2m} = 1$ whenever p_{2m} is defined, or $p_i = 0$ or 1 for some i, $1 \leq i \leq 2m-1$.

The next theorem gives a sufficient and a necessary condition for admissibility.

Theorem 3.2.2: Let $S(\xi)$ denote the spectrum of ξ.

(i) If there exists a positive definite matrix T such that $(f(x), Tf(x)) \leq 1$ for all $x \in [a,b]$ and equality holds for $x \in S(\xi)$ then ξ is admissible. (The notation (u,v) signifies the inner product of the vectors u and v.)

(ii) If ξ is admissible then there exists a nonnegative matrix T (not necessarily positive definite) such that $(f(x), Tf(x)) \leq 1$ for all $x \in [a,b]$ with equality occurring for $x \in S(\xi)$.

Proof: See Karlin and Studden (1966b).

The following theorems give necessary conditions for admissibility for several weight functions.

Theorem 3.2.3: Let $f^T(x) = \sqrt{w(x)}\,(1,x,\ldots,x^m)$ where $w(x) = [x(1-x)]^{\alpha+1}|x - \frac{1}{2}|^{\gamma}$, $x \in [0,1]$, $\alpha \geq -1$, $\gamma > 0$. Suppose ξ is admissible. If m is odd then the spectrum of ξ contains at most m+1 points. If m is even then the spectrum of ξ contains at most m+2 points.

Proof: Assume $\alpha > -1$. If ξ is admissible, then Theorem 3.2.2 implies the existence of a nonnegative polynomial $T_{2m}(x)$ of degree at most 2m such that

$$w(x)P_{2m}(x) \le 1$$

for all $x \in [0,1]$ and equality holds on the spectrum of ξ. Suppose the spectrum of ξ contains r points. Since $[x(1-x)]^{\alpha+1}|x - \frac{1}{2}|^{\gamma}$ vanishes at x = 0 and 1 the equation

$$L(x) = w(x)P_{2m}(x)-1 = 0$$

has at least 2r zeros, counting multiplicities. The derivative of $w(x)P_{2m}(x)$ is found to be

$$\begin{cases} [x(1-x)]^{\alpha}|x - \frac{1}{2}|^{\gamma}Q_{2m+2}(x) & x \in (0, \frac{1}{2}) \\ [x(1-x)]^{\alpha}|x - \frac{1}{2}|^{\gamma}Q_{2m+2}(x) & x \in (\frac{1}{2}, 1) \end{cases}$$

where $Q_{2m+2}(x)$ is a polynomial of degree 2m+2. By symmetry, L(x) = 0 has r zeros on each region. Thus, by Rolles' Theorem, L'(x) has at least 2r-2 zeros other than 0,1 and $\frac{1}{2}$. On the other hand, L'(x) has at most 2m+2 zeros other than 0,1 and $\frac{1}{2}$. It follows that $2r-2 \le 2m+2$ or $r \le m+2$. In case m is odd, one can see that the design with m+2 points must have 0 in its support which is absurd. So r = m+1 follows. The proof for the case $\alpha = -1$ is similar.

<u>Theorem 3.2.4</u>: ξ is admissible for $f(x) = \sqrt{w(x)}\,(1,x,\dots,x^m)$, $x \in [0,1]$ implies that the support of ξ has no more than m+1 points, where w(x) is one of the following:

(1) $w(x) = x^{\alpha+1}(1-x)^{\beta+1}$, $\alpha > -1$, $\beta > -1$.

(2) $w(x) = x^{\alpha+1}$, $\alpha > -1$.

(3) $w(x) = (1-x)^{\beta+1}$, $\beta > -1$.

(4) $w(x) = (1-|x - \frac{1}{2}|)^{k}$, $k > 0$.

(5) $w(x) = 1-\sqrt{x}$.

(6) $w(x) = \sqrt{x}(1-\sqrt{x})$.

(7) $w(x) = x(1-\sqrt{x})$.

Proof: Karlin and Studden (1966a) gave the proof of (1). The others can be proved similarly.

3.3 The Weighted D-Optimal Design for Polynomial Regression

The D-optimal design ($w(x) \equiv 1$) was first solved by Guest (1958). Studden (1980) gave a new proof using the method of canonical moments.

Theorem 3.3.1. The D-optimal design for $f^T(x) = (1,x,\ldots,x^m)$ has canonical moments

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,m-1$$

$$p_{2i} = \frac{m-i+1}{2m-2i+1} \qquad i = 1,2,\ldots,m-1$$

$$p_{2m} = 1.$$

It puts weight $\frac{1}{m+1}$ on each of the zeros of $x(x-1)P_{m-1}^{(1,1)}(x) = 0$, where $\{P_k^{(1,1)}(x)\}$ are orthogonal with respect to $w(x) = x(1-x)$.

Proof: See Studden (1980).

The method of canonical moments was applied to the problem of weighted D-optimal design.

Theorem 3.3.2: Let $f^T(x) = \sqrt{w(x)}\,(1,x,\ldots,x^m)$ where $w(x)$ is one of the following:

(1) $w(x) = x^{\alpha+1}(1-x)^{\beta+1}$, $\alpha > -1$, $\beta > -1$

(2) $w(x) = x^{\alpha+1}$, $\alpha > -1$

(3) $w(x) = (1-x)^{\beta+1}$, $\beta > -1$.

Then the determinant $|M(\xi)|$ is uniquely maximized by the design ξ having canonical moments

$$p_{2i+1} = \frac{\alpha+m+1-i}{\alpha+\beta+2(m+1-i)} \qquad i = 0,1,\dots,m$$

$$p_{2i} = \frac{m+1-i}{\alpha+\beta+1+2(m-i+1)} \qquad i = 1,2,\dots,m+1.$$

For case (1), we have $p_{2m+2} = 0$. The design ξ puts equal mass $\frac{1}{m+1}$ on each of the zeros of $P_{m+1}^{(\alpha,\beta)}(x) = 0$. For case (2), we have $p_{2m+1} = 1$. The design ξ puts equal mass $\frac{1}{m+1}$ on each of the zeros of $(x-1)P_m^{(\alpha,1)}(x) = 0$. For case (3), we have $p_{2m+1} = 0$. The design ξ puts equal mass $\frac{1}{m+1}$ on each of the zeros of $xP_m^{(1,\beta)}(x) = 0$.

Proof: See Studden (1982).

Theorem 3.3.3: Let $f^T(x) = \sqrt{w(x)}\,(1,x,\dots,x^m)$ where

$w(x) = [x(1-x]^{\alpha+1}|x - \frac{1}{2}|^{\gamma}$, $\alpha \geq -1$, $\gamma > 0$, $x \in [0,1]$.

If m is odd, the determinant $|M(\xi)|$ is maximized by the measure ξ^* having canonical moments

$$p_{2i+1} = \frac{1}{2}, \qquad i = 0,1,\dots,m$$

$$p_{2i} = \begin{cases} \dfrac{\gamma+m-i+1}{2\alpha+\gamma+2(m-i+1)} & i = 1,3,\dots,m \\[2ex] \dfrac{m-i+1}{2\alpha+\gamma+1+2(m-i+1)} & i = 2,4,\dots,m+1. \end{cases}$$

If $\alpha > -1$, then $P_{2m+2} = 0$. The weighted D-optimal design ξ^* puts equal mass $\frac{1}{m+1}$ on each of the zeros of $P_{m+1}(x) = 0$ where $\{P_k(x)\}$ are orthogonal with respect to $w(x) = [x(1-x)]^{\alpha}|x - \frac{1}{2}|^{\gamma}$. If $\alpha = -1$, then $P_{2m} = 1$. The weighted D-optimal design ξ^* puts equal mass $\frac{1}{m+1}$ on each of the zeros of $x(x-1)P_{m-1}(x) = 0$ where $\{P_k(x)\}$ are orthogonal with respect to $w(x) = x(1-x)|x - \frac{1}{2}|^{\gamma}$.

Proof: Notice that we can consider symmetric designs only since $w(x)$ is symmetric about $\frac{1}{2}$. By Theorem 3.2.3, we see that the support of the weighted D-optimal design consists of m+1 points, say x_0, x_1, $\ldots, x_m$. Then $|M(\xi)|$ can be written

$$|M(\xi)| = \prod_{i=0}^{m} w(x_i) \quad |M_0(\xi)|$$

where $M_0(\xi)$ is the information matrix when $w(x) \equiv 1$. By Theorem 2.4.10, we have

$$\prod_{i=0}^{m} x_i = \zeta_1\zeta_3 \cdots \zeta_{2m+1} = (\tfrac{1}{2})^{m+1} q_2 q_4 \cdots q_{2m}$$

$$\prod_{i=0}^{m} (1-x_1) = \prod_{i=1}^{2m+1} q_i = (\tfrac{1}{2})^{m+1} q_2 q_4 \cdots q_{2m}$$

and
$$\prod_{i=0}^{m} |x_i - \tfrac{1}{2}| = (\tfrac{1}{4})^{2m} p_2 q_4 p_6 \cdots p_{2m}.$$

So $|M(\xi)|$ can be written

$$\text{const}\Big(\prod_{i=1}^{m} q_{2i}^2\Big)^{\alpha+1} (p_2 q_4 p_6 \cdots p_{2m})^{\gamma} \prod_{i=2}^{m} (q_{2i-2} p_{2i})^{m-i+1} p_2^m.$$

By rearranging the terms, we obtain the desired results. Comparing the p_i's we obtain here with those of Example II.4.1, we find the support of ξ^* is the zeros of $P_{m+1}(x) = 0$ where $\{P_k(x)\}$ are orthogonal with respect to $w(x) = [x(1-x)]^{\alpha}|x - \frac{1}{2}|^{\gamma}$ if $\alpha > -1$. If $\alpha = -1$, $\{P_k(x)\}$ are orthogonal with respect to $w(x) = x(1-x)|x - \frac{1}{2}|^{\gamma}$. The proof is completed.

The problem we have solved in the last theorem can be formulated in another way. Let us introduce some notations and preliminary results. Let $\mathcal{J}_{m+1}$ denote the set of (m+1)-tuples $(x_0, x_1, \ldots, x_m)$ with $0 \leq x_0 < x_1 < \ldots < x_m \leq 1$. Further let us define

$$r_v(x) = \frac{1}{\sqrt{w(x_v)}} \ell_v(x), \qquad v = 0,1,\ldots,m$$

where $$\ell_v(x) = \frac{\ell(x)}{\ell'(x_v)(x-x_v)}, \qquad v = 0,1,\ldots,m$$

and $$\ell(x) \quad \prod_{i=0}^{m} (x-x_i).$$

The following lemma has been proved by Karlin and Studden (1966a).

<u>Lemma 3.3.4</u>: Let $f^T(x) = \sqrt{w(x)}\,(1, x, \ldots, x^m)$. Then we have

$$d(x,\xi) = (m+1)w(x) \sum_{i=0}^{m} r_i^2(x).$$

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

The problem of interest is to find a suitable choice of x_0, $x_1, \ldots, x_m$ so that

$$\max w(x) \sum_{i=0}^{m} r_i^2(x)$$

is minimized. By Theorem 3.1.1, the equivalence theorem, we see that

$$\max w(x) \sum_{i=0}^{m} r_i^2(x) \geq 1$$

and the equality sign holds when $x_0,\ldots,x_m$ are elements of the spectrum of the weighted D-optimal design. We thus prove the following theorem.

Theorem 3.3.4: Let m be an odd natural number. If $w(x) = [x(1-x)]^{\alpha+1}|x - \frac{1}{2}|^{\gamma}$, where $\alpha \geq -1$ and $\gamma > 0$, is defined on [0,1], then

$$\inf_{\mathcal{J}_{m+1}} \sup_{0\leq x\leq 1} [x(1-x)]^{\alpha+1}|x - \tfrac{1}{2}|^{\gamma}\{r_0^2(x) +\ldots+ r_m^2(x)] = 1$$

and the infimum is uniquely attained when $x_0,x_1,\ldots,x_m$ are the zeros of $P_{m+1}(x) = 0$, where $\{P_k(x)\}$ are orthogonal with respect to $w(x) = [x(1-x)]^{\alpha}|x - \frac{1}{2}|^{\gamma}$ if $\alpha > -1$. If $\alpha = -1$ then $\{P_k(x)\}$ are orthogonal with respect to $w(x) = x(1-x)|x - \frac{1}{2}|^{\gamma}$.

Remark 3.3.4: The case $w(x) = x^{\alpha+1}(1-x)^{\beta+1}$, $\alpha > -1$, $\beta > -1$ has been already solved by Karlin and Studden (1966a).

The case where m is even seems intractable. The following example gives the weighted D-optimal design when $w(x) = |x - \frac{1}{2}|^2$ and m = 2.

Example 3.3.1: Let x = [0,1], $w(x) = (x - \frac{1}{2})^2$ and $f^T(x) = (1,x,x^2)$. Let us transform the interval to [-1,1]. The weight function becomes $x^2, x \in [-1,1]$. The information matrix $M(\xi)$ can be written

$$\begin{bmatrix} \mu_2 & 0 & \mu_4 \\ 0 & \mu_4 & 0 \\ \mu_4 & 0 & \mu_6 \end{bmatrix}$$

Let $y = x^2$. The determinant of the above matrix can be written

$$|\mu_2'| \begin{vmatrix} \mu_1' & \mu_2' \\ \mu_2' & \mu_3' \end{vmatrix}$$

where μ_i' is the i-th moment of $y(=x^2)$. The determinant can be written in terms of canonical moments

$$p_1(p_1 + q_1p_2)p_1^2 \, q_1p_2q_2p_3.$$

After a tedious calculation we find the above expression is maximized if $p_1 = \frac{31-\sqrt{61}}{30}$, $p_2 = \frac{11-\sqrt{61}}{6}$ and $p_3 = 1$. Converting back to the interval [-1,1], we have

$$p_{2i+1} = \frac{1}{2}, \qquad i = 0,1,2$$

$$p_2 = \frac{31-\sqrt{61}}{30}$$

$$p_4 = \frac{11-\sqrt{61}}{6}$$

$$p_6 = 1.$$

The design is supported on 4 points. The support is the zeros of $x(x-1)Q_2(x) = 0$, where $\{Q_k(x)\}$ are orthogonal with respect to $x(1-x)d\xi$. In this case $Q_2(x) = x^2 - \frac{1}{2}\,\frac{\sqrt{61}+1}{6}\,x + \frac{11-\sqrt{61}}{120} = 0$.

3.4 Regression Function through the Endpoints or Midpoint of the Regression Interval

Sometimes the regression function will assume some specified values at some points. For example, we want to find the relationship between the speed (x) and the distance (y) to stop for a car. Suppose the model is a polynomial of degree m, say $E(y) = \beta_0 + \beta_1 x + \ldots + \beta_m x^m$. Without taking any data we are quite sure that $y = 0$ if $x = 0$. This implies $\beta_0 = 0$ or some other lower coefficients may vanish also, say $\beta_0 = \beta_1 = \ldots = \beta_r = 0$. The model can then be written

$$E(y) = \beta_{r+1} x^{r+1} + \beta_{r+2} x^{r+2} + \ldots + \beta_m x^m.$$

We recognize at once that this model is the weighted regression model with weight function $w(x) = x^{2r+2}$. Sometimes the regression function may assume some specified value y_0 at some point $x_0 \in [-1,1]$. Then the model can be written

$$E(y) = y_0 + \beta_1(x-x_0) + \ldots + \beta_m(x-x_0)^m.$$

In case $\beta_1 = \ldots = \beta_r = 0$, we have

$$E(y) = y_0 + \beta_{r+1}(x-x_0)^{r+1} + \ldots + \beta_m(x-x_0)^m.$$

If we make a transformation $z = y-y_0$, we see the model becomes

$$E(z) = \beta_{r+1}(x-x_0)^{r+1} + \ldots + \beta_m(x-x_0)^m.$$

It is readily seen that the above model is the same as the weighted regression model with weight function $w(x) = (x-x_0)^{2r+2}$.

In particular x_0 may be an end-point or the midpoint of the regression interval. The D-optimal designs for the resulting models have been found in preceding sections.

If the regression function vanishes at both endpoints and the midpoint of the regression interval, then the regression function becomes, by factorization,

$$E(y) = x^{\alpha}(1-x)^{\beta}(x - \tfrac{1}{2})^{\gamma}(\beta_0 + \beta_1 x + \ldots + \beta_{m-\alpha-\beta-\gamma} x^{m-\alpha-\beta-\gamma})$$

where α, β, γ are non-negative integers and $m-\alpha-\beta-\delta \geq 0$. So again we go back to the weighted regression model with weight function $w(x) = x^{2\alpha}(1-x)^{2\beta}(x - \frac{1}{2})^{2\gamma}$.

3.5 Trigonometric Regression

In this section we will discuss the D-optimal design for the complete set of trigonometric functions, namely, $f^T(x)=(1,\cos x,\sin x,\ldots,\cos mx,\sin mx)$ $x \in [0,2\pi]$, and the admissibility problem. Karlin and Studden (1966a) proved that the D-optimal design for $f^T(x) = (1, \cos x, \cos 2x,\ldots,\cos mx)$ on $[0,\pi]$ can be reduced to the D-optimal design for polynomial regression of degree m on $[-1,1]$. It should be noted that the same conclusion holds for $f^T(x) = (1, \sin x, \sin 2x,\ldots,\sin mx)$ on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ since $\sin(x-\frac{\pi}{2}) = \cos x$. The same authors proved that the D-optimal design with the smallest size of support for the complete set of trigonometric functions are those measures concentrating mass $\frac{1}{2m+1}$ on each of the 2m+1 equidistant points on the circle. Here we will use the parameters that we introduced in Chapter II section 6 to characterize all measures that will maximize $|M(\xi)|$. The proof is new and the spirit of this approach can be extended to other problems.

Let us write $f^T(x) = (1, 2\cos x, 2\sin x,\ldots,2\cos mx, 2\sin mx)$. Notice that we put 2 before the sine and cosine functions for the sake of convenience and it will not affect the D-optimal design. By Euler's formula,

$$2\cos kx = e^{ikx} + e^{-ikx} \quad \text{and}$$

$$2\sin kx = ie^{-ikx} - ie^{ikx},$$

we can write

$$f(x) = S\, g(x)$$

where

$$S = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & i & 0 & -i & \cdots & 0 & 0 \\ 1 & 0 & \cdots & & & & & 0 & 1 \\ i & 0 & \cdots & & & & & 0 & -i \end{pmatrix}, \qquad (3.5.1)$$

$$g^T(x) = (e^{-imx}, e^{-i(m-1)x}, \ldots, 1, \ldots, e^{i(m-1)x}, e^{imx}).$$

Since S is a linear transformation, we have

$$f^T(x)M_f^{-1}(\xi)f(x) = g^T(x)M_g^{-1}(\xi)g(x), \qquad (3.5.2)$$

where $M_f(\xi) = \int_0^{2\pi} f(x)f^T(x)d\xi(x)$ and

$$M_f(\xi) = \int_0^{2\pi} g(x)g^T(x)d\xi(x).$$

Note that we put the regression function f(x) as the index of the information matrix to emphasize the dependence of the information matrix on the regression function. Let $c_n = \int_0^{2\pi} e^{-inx}d\xi(x)$. We can write $M_g(\xi)$ as

$$\begin{pmatrix} c_{2m} & c_{2m-1} & \cdots & c_1 & 1 \\ c_{2m-1} & c_{2m-2} & \cdots & 1 & c-1 \\ \vdots & \vdots & & & \\ c_1 & 1 & \cdots & -c_{-2m+2} & c_{-2m+1} \\ 1 & c_{-1} & \cdots & c_{-2m+1} & c_{-2m} \end{pmatrix}.$$

It can be easily seen that

$$S^T M_f^{-1}(\xi) S = M_g^{-1}(\xi)$$

where S is defined in (3.5.1). So we have

$$|M_f(\xi)| = |M_g(\xi)|\ |s|^2$$
$$= (2i)^{2m}\ |M_g(\xi)|.$$

Let

$$J = \begin{pmatrix} 0 & & & 1 \\ & & 1 & \\ & \cdot^{\cdot^{\cdot}} & & \\ 1 & & & 0 \end{pmatrix}.$$

Recall that the Toeplitz matrix of size 2m+1 is given by

$$T_{2m}(\xi) = \begin{pmatrix} 1 & c_1 & \cdots & c_{2m-1} & c_{2m} \\ c_{-1} & 1 & \cdots & c_{2m-2} & c_{2m-1} \\ \vdots & & & \vdots & \vdots \\ c_{-2m+1} & c_{-2m+2} & \cdots & 1 & c_{-1} \\ c_{-2m} & c_{-2m+1} & \cdots & c_{-1} & 1 \end{pmatrix}$$

and the determinant of $T_{2m}(\xi)$ is denoted by Δ_{2m}. It is clear that $M_g(\xi) = T_{2m}(\xi)J$ and thus $|M_g(\xi)| = (-1)^m\Delta_{2m}$. So the maximization of $|M_f(\xi)|$ is equivalent to the maximization of Δ_{2m}. By (2.6.8), Δ_{2m} can be expressed in terms of the parameters $\{a_k\}$, namely

$$\Delta_{2m} = (1-|a_0|^2)^{2m}(1-|a_1|^2)^{2m-1}\ldots(1-|a_{2m-1}|^2)$$

where $|a_k| \leq 1$, $k = 0,1,\ldots,2m-1$. It is clear now that Δ_{2m} is maximized iff

$$a_0 = a_1 = \ldots = a_{2m-1} = 0.$$

So we have the following theorem.

Theorem 3.5.1: Let $f^T(x) = (1, 2\cos x, 2\sin x, \ldots, 2\cos mx, 2\sin mx)$, $x \in [0,2\pi]$, then $|M_f(\xi)|$ is maximized by those measures having

$$a_0 = a_1 = \ldots = a_{2m-1} = 0.$$

There are infinitely many measures representing $a_0 = a_1 = \ldots = a_{2m-1} = 0$. One possible choice is the continuous uniform measure on the circle. In Example 2.6.3, we showed that the continuous uniform distribution is characterized by $a_i = 0$ for all i and so it is a possible candidate. Another possible choice is those measures that put mass $\frac{1}{2m+1}$ on each of well defined 2m+1 equidistant points on the circle. In Example 2.6.2, these measures are characterized by

$$\begin{aligned} & a_0 = a_1 = \ldots = a_{2m-1} = 0 \\ \text{and} \quad & |a_{2m}| = 1 \end{aligned} \qquad (3.5.3)$$

Similarly, those measures that put equal mass $\frac{1}{2m+1+n}$, on each of 2m+1+n equidistant points (where n is a natural number) are also D-optimal. On the other hand, the measures characterized by (3.5.3) are also D-optimal for trigonometric regression of degree less than m.

In the following we will find the variance function of those designs characterized by (3.5.3). Since

$$M_g(\xi) = T_{2m}(\xi)J,$$

by noting that $g(x)^T J^T = g(x)^T J = \overline{g(x)}$, we have

$$g(x)^T M_g^{-1}(\xi) g(x) = \overline{g(x)}^T T_{2m}^{-1}(\xi) g(x), \qquad (3.5.4)$$

where $\overline{g(x)}$ denotes the complex conjugate of g(x). Since $a_0 = a_1 = \ldots = a_{2m-1} = 0$, T_{2m} is reduced to the identity matrix. By (3.5.2) and (3.5.4), we see that

$$\begin{aligned} d(x,\xi) &= f(x)^T M_f^{-1}(\xi) f(x) \\ &= g(x)^T M_g^{-1}(\xi) g(x) \end{aligned}$$

$$= \overline{g(x)}^T T_{2m}^{-1}(\xi) g(x)$$

$$= \overline{g(x)}^T g(x).$$

Noting that

$$g^T(x) = (e^{-imx},\ldots,1,\ldots,e^{imx}),$$

we see that $\overline{g(x)}^T g(x) = 2m+1$. As a consequence, the confidence interval for the trigonometric regression has a constant width.

In the following we will discuss the admissibility problem for the trigonometric regression. Recall that ξ is admissible iff there does not exist a design ξ' such that $M(\xi') - M(\xi)$ is positive semi-definite and $M(\xi') \neq M(\xi)$. Since $M_f(\xi) = SM_g(\xi)S^T$ and $M_g(\xi) = T_{2m}(\xi)J$, we have $M_f(\xi) = ST_{2m}(\xi)JS^T$. Let $V = JS^T$. It is readily checked that $S = \bar{V}^T$, so we have $M(\xi) = \bar{V}^T T_{2m}(\xi)V$. Since $T_{2m}(\xi)$ is a Hermitian matrix and V is non-singular, we see that $M(\xi)$ is positive (semi-) definite iff $T_{2m}(\xi)$ is positive (semi-) definite. It follows that the admissibility for the trigonometric regression can be written in terms of the Toeplitz matrix: ξ is admissible iff there does not exist a design ξ' such that $T_{2m}(\xi') - T_{2m}(\xi)$ is positive semi-definite and $T_{2m}(\xi') \neq T_{2m}(\xi)$. We have the following theorem.

Theorem 3.5.2: Every design is admissible for the trigonometric regression.

Proof: Suppose ξ is not admissible. There exists ξ' such that $T_{2m}(\xi') - T_{2m}(\xi)$ is positive semi-definite and $T_{2m}(\xi) \neq T_{2m}(\xi')$. Notice that both ξ and ξ' are probability measures, the diagonal elements of both $T_{2m}(\xi)$ and $T_{2m}(\xi')$ are identically one. Thus the diagonal elements of $T_{2m}(\xi) - T_{2m}(\xi')$ are identically zero. The positive semi-definiteness of $T_{2m}(\xi) - T_{2m}(\xi')$ implies that $T_{2m}(\xi') \equiv T_{2m}(\xi)$. A contradiction.

3.6 D-Optimal Rotatable Designs

Let X denote the unit m-dimensional sphere and let $\{f_i: \ i = 0,1, \ldots,n\}$ be the set of all functions of the form $\prod_{i=1}^{m} x_i^{\alpha_i}$, where α_j's are nonnegative integers satisfying $\sum_{j=1}^{m} \alpha_j \le d$ for some positive integers d. It is known that n can be determined from m and d, namely, $n = \binom{m+d}{m}$. We want to find a design ξ^* such that $|M(\xi)|$ is maximized among the orthogonal invariant measures - the measures that are invariant under orthogonal transformation on X. It is also known that any orthogonal invariant measure ξ can be written as

$$\xi(A) = \int_0^1 \xi_1(r^{-1}(A \cap S_r))\rho(dr)$$

where S_r is the (m-1)-sphere of radius r, ξ_1 is the uniform probability measure on S_1, ρ is a probability measure on [0,1], and the integrand, when r = 0, is taken to be one if $0 \in A$ and zero if $0 \notin A$. So ξ can be decomposed into two parts, one is the radial component, ρ, and the other is the uniform measure on the (m-1)-sphere, ξ_1. Our aim here is to investigate the radial component of the D-optimal rotatable design.

The entries of $M(\xi)$ are given by

$$\int x_1^{\alpha_1} \ldots x_m^{\alpha_m} x_1^{\beta_1} \ldots x_m^{\beta_m} d\xi(x) = c(\alpha_1+\beta_1,\ldots,\alpha_m+\beta_m)\mu_{\Sigma(\alpha_1+\beta_1)}$$

where $c(\alpha_1+\beta_1,\ldots,\alpha_m+\beta_m)$ denotes the $(\alpha_1+\beta_1,\ldots,\alpha_m+\beta_m)$th moment of the uniform measure on the (m-1)-sphere and μ_s is the s-moment of ρ. Notice that

$c(\alpha_1+\beta_1,\ldots,\alpha_m+\beta_m) = 0$ if there exists one i such that $\alpha_i + \beta_i$ is odd. For more details, see Kiefer (1961b) or Karlin and Studden (1966a). The moments of the uniform measure on the (m-1)-sphere can be found by differentiating the Fourier transform of ξ. The following result is given by Shilov (1965).

Lemma 3.6.1: Let $F(\sigma)$ denote the Fourier transform of the uniform measure on the m-1 sphere. We have

$$F(\sigma) = \frac{2^{\frac{m}{2}-1}\Gamma(\frac{m}{2})}{(\sigma)^{\frac{m}{2}-1}} J_{\frac{m}{2}-1}(|\sigma|) \tag{3.6.1}$$

where $|\sigma| = \sqrt{\sigma_1^2+\ldots+\sigma_m^2}$ and $J_{\frac{m}{2}\;1}(x)$ is the Bessel function of order $\frac{m}{2}-1$.

Proof: See Shilov (1965).

By expanding the Bessel function $J_{\frac{m}{2}-1}(x)$ in power series, (3.6.1) can be written

$$\Gamma(\tfrac{m}{2}) \sum_{j=0}^{\infty} \frac{(-1)^j|\sigma|^{2j}}{j!2^{2j}\Gamma(\frac{m}{2}+j)}$$

$$= 1 - \frac{\Gamma(\frac{m}{2})|\sigma|^2}{2^2\Gamma(\frac{m}{2}+1)} + \frac{\Gamma(\frac{m}{2})|\sigma|^4}{2!2^4\Gamma(\frac{m}{2}+2)} - \ldots$$

So, by partial differentiation, we have

$$c(2,0,\ldots,0) = c(0,2,0,\ldots,0) = \ldots = c(0,\ldots,0,2) = \frac{1}{m}$$

$$c(4,0,\ldots,0) = c(0,4,0,\ldots,0) = \ldots = c(0,\ldots,0,4) = \frac{3}{m(m+2)}$$

$$c(2,2,0,\ldots,0) = c(2,0,2,0,\ldots,0) = \ldots = \frac{1}{m(m+2)}$$

$$c(6,0,\ldots,0) = c(0,6,0,\ldots) = \ldots = \frac{15}{m(m+2)(m+4)}$$

$$c(4,2,0,\ldots,0) = c(2,4,0,\ldots,0) = \ldots = \frac{3}{m(m+2)(m+4)}$$

In the following example, we will show how to find ρ by using canonical moments.

Example 3.6.1: $(d = 1)$. It is known that

$$c(2,0,\ldots,0) = c(0,2,0,\ldots,0) = \ldots = c(0,\ldots,0,2) = \frac{1}{m}.$$

So we have

$$M(\xi) = \begin{bmatrix} 1 & & & \\ & \frac{1}{m}\mu_2 & & 0 \\ & & \ddots & \\ 0 & & & \frac{1}{m}\mu_2 \end{bmatrix}$$

where $\mu_2 = \int_0^1 r^2 d\rho(r)$.

If we make the transformation $t = r^2$, then

$$\mu_2 = \int_0^1 t \, d\rho'(t) = \mu_1'$$

where $\rho'(t) = \rho(r^2)$ and μ_1' is the first moment of ρ'. Hence we have

$$|M(\xi)| = \left(\frac{1}{m}\mu_1'\right)^m = \left(\frac{1}{m}p_1\right)^m.$$

Here p_k is the canonical moment of ρ'. It is clear that $p_1 = 1$, i.e. we have to put all the mass on the surface of the (m-1)-sphere.

Example 3.6.2: $(d = 2)$. Write $f^T(x) = (1, x_1^2, \ldots, x_m^2, x_1, \ldots, x_m, x_1x_2, \ldots, x_{m-1}x_m)$ and we find

$$M(\xi) = \begin{bmatrix} 1 & \frac{1}{m}\mu_2 & \cdots & \frac{1}{m}\mu_2 & & & & & \\ \frac{1}{m}\mu_2 & \frac{3}{m(m+2)}\mu_4 & \cdots & \frac{1}{m(m+2)}\mu_4 & & & & & \\ & & \frac{3}{m(m+2)}\mu_4 & & & & & & \\ & \vdots & & \vdots & & & & & \\ & \frac{1}{m(m+2)}\mu_4 & \cdots & \frac{3}{m(m+2)}\mu_4 & & & & & \\ & & & & \frac{1}{m}\mu_2 & & & & \\ & & & & & \ddots & & & \\ & & & & & & \frac{1}{m}\mu_2 & & \\ & & & & & & & \frac{1}{m(m+2)}\mu_4 & \\ & & & & & & & & \ddots \\ & & & & & & & & & \frac{1}{m(m+2)}\mu_4 \end{bmatrix}.$$

Immediately, we have

$$|M(\xi)| = \text{const.}\ (\mu_4-\mu_2^2)\mu_2^m\mu_4^{\frac{(m-1)(m+2)}{2}}.$$

Using the transformation $t = r^2$, we have

$$|M(\xi)| = \text{const.}\ (\mu_2'-\mu_1'^2)(\mu_1')^m\mu_2'^{\frac{(m-1)(m+2)}{2}}$$

$$= \text{const.}\ p_1^{1+m+\frac{(m-1)(m+2)}{2}}q(p_1+q_1p_2)^{\frac{(m-1)(m+2)}{2}}.$$

By differentiating the above expression, we have

$$p_1 = \frac{m(m+3)}{(m+1)(m+2)}$$

$$p_2 = 1.$$

By direct calculation, we see the resulting design puts mass $\frac{2}{(m+1)(m+2)}$ $(= q_1p_2)$ at the centre and $1-\frac{2}{(m+1)(m+2)} = \frac{m(m+3)}{(m+1)(m+2)}$ on the surface on the (m-1) sphere. The computation for $d \geq 3$ is difficult.

CHAPTER IV

D_s-OPTIMAL DESIGNS

4.1 General Results

Suppose we are interested in only s out of the m+1 parameters in the regression model

$$E(y) = \sqrt{w(x)}\ (\beta_0 + \ldots + \beta_m x^m) = \beta^T f(x)$$

where $x \in [a,b]$ and $w(x)$ is a known nonnegative continuous function. Let us write $f(x)$ in the form $\binom{f_1(x)}{f_2(x)}$ so that $f_2(x)$ is the s-vector that corresponds to the s parameters of interest and $f_1(x)$ is a (r+1)-vector (r = m-s). Similarly we can write $M(\xi)$ as

$$\begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) \end{bmatrix} .$$

It is known that

$$|\Sigma_s(\xi)| = \frac{|M(\xi)|}{|M_{11}(\xi)|} = |M_{22}(\xi) - M_{21}(\xi) M_{11}^{-1}(\xi) M_{12}(\xi)|$$

is proportional to the inverse of the generalized variance of the least squares estimator of the s-vector of coefficients. The D_s-optimal design is defined to be the design that maximizes $|\Sigma_s(\xi)|$. The following is an equivalence theorem for D_s-optimal designs.

Theorem 4.1.1: If $|M(\xi^*)| \neq 0$, the following assertions are equivalent.

(1) The design ξ^* maximizes $|\Sigma_s(\xi)|$

(2) The design ξ^* minimizes $\max_x d_s(x,\xi)$, where

$$d_s(x,\xi) = f^T(x)M^{-1}(\xi)f(x) - f_1^T(x)M_{11}^{-1}(\xi)f_1(x).$$

(3) $\max_x d_s(x,\xi^*) = s$.

Proof: See Kiefer (1961a).

In most of the following sections we will concentrate on the discussion of the D_s-optimal design for the s highest coefficients of the regression polynomial. We have the following theorem.

Theorem 4.1.2: Let $x^* = ax+b$, $a > 0$, be a transformation from $[0,1]$ to $[b, a+b]$. Suppose $w(x)$ undergoes the same transformation. Let ξ be the D_s-optimal design for the s highest coefficients on $[0,1]$. If ξ' is the design induced by ξ on $[b, a+b]$, then ξ' is the D_s-optimal design for the s highest coefficients on $[b, a+b]$.

Proof: Let $g^T(x^*) = \sqrt{w(x^*)}(1, x^*, \ldots, x^{*m})$ and $x^* \in [b,a+b]$. Decompose $g(x^*)$ into two components as we did for $f(x)$, i.e.

$$g^T(x^*) = (g_1^T(x^*), g_2(x^*))$$

where $g_1(x^*)$ is a $(r+1)$-vector, and $g_2(x^*)$ is a s-vector $(r+s=m)$. We see that

$$g(x^*) = \begin{bmatrix} g_1(x^*) \\ g_2(x^*) \end{bmatrix} = L\ f(x) = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} .$$

The result follows at once.

4.2 The D_s-Optimal Design in Case $w(x) = 1$

The following result was given by Studden (1981). We will give a shorter proof here.

Theorem 4.2.1: Let $w(x) = 1$. The D_s-optimal design ξ^* is given by

$$p_i = \frac{1}{2} \qquad i \text{ is odd}$$

$$p_{2i} = \begin{cases} \frac{1}{2} & i = 1,2,\ldots,r \\ \frac{m-i+1}{2(m-i)+1} & i = r+1,\ldots,m-1 \end{cases}$$

$$p_{2m} = 1.$$

The support of ξ^* is given by the zeros of

$$x(x-1)[p_r^{(\frac{1}{2},\frac{1}{2})}(x)p_{s-1}^{(1,1)}(x) - \frac{1}{8}\,\frac{s-1}{2s-1}\,p_{r-1}^{(\frac{1}{2},\frac{1}{2})}(x)p_{s-2}^{(1,1)}(x)] \qquad (4.2.1)$$

where $\{p^{(i,j)}(x)\}$ denotes the monic polynomials orthogonal with respect to $x^i(1-x)^j$. If we order the zeros of (4.2.1) in such a way so that $0 = x_1 < x_2 <\ldots< x_{m+1} = 1$, and define $2x_i-1 = \cos\theta_i$, then the weight attached to each point x_i is given by

$$\xi^*(x_i) = \frac{2}{2m+1+\dfrac{\sin(2r+1)\theta_i}{\sin\theta_i}}\,. \qquad (4.2.2)$$

Proof: Here we give a new proof of (4.2.1) only. The rest of the proof can be found in Studden (1981). By Theorem 2.6.1, the support of ξ^* is given by the zeros of

$$K\begin{pmatrix} -\frac{1}{2} & -\frac{1}{4} & \cdots & -\frac{1}{4} & -\frac{1}{2}\frac{s-1}{2s-1}, & \ldots, & -\frac{2}{5}\frac{1}{2} & -\frac{1}{2}\frac{2}{3} & -\frac{1}{3}\frac{1}{2} & -\frac{1}{2} \\ x & 1 & x \ldots x & 1 & & x \ldots x & 1 & x & 1 & x \end{pmatrix}.$$

Using Lemma 2.7.3, we can write this as

$$x(x-1)K\begin{pmatrix} -\frac{1}{4} & \cdots & -\frac{1}{4} & -\frac{1}{2}\frac{s-1}{2s-1} & -\frac{1}{2}\frac{s}{2s-1} & \cdots & -\frac{3}{5}\frac{1}{2} & -\frac{1}{2}\frac{1}{3} & -\frac{2}{3}\frac{1}{2} & \\ 1 & x \ldots x & 1 & x & & 1 \ldots x & 1 & x & 1 \end{pmatrix}.$$

Making an odd contraction, we have

$$x(x-1)K\begin{pmatrix} -\frac{1}{16} & \cdots & -\frac{1}{16} & -\frac{1}{8}\frac{s-1}{2s-1} & -\frac{1}{4}\frac{s(s-2)}{(2s-1)(2s-3)} & \cdots & -\frac{1}{4}\frac{1}{5} \\ x-\frac{1}{2} & x-\frac{1}{2}\ldots & x-\frac{1}{2} & x-\frac{1}{2} & & \cdots & x-\frac{1}{2} \end{pmatrix}.$$

By Theorem 2.3.2, we obtain

$$x(x-1)\left[K\begin{pmatrix} -\frac{1}{16} & \cdots & -\frac{1}{16} \\ x-\frac{1}{2} & \cdots & x-\frac{1}{2} \end{pmatrix} K\begin{pmatrix} -\frac{1}{4}\frac{s-2}{(2s-1)(2s-3)} & \cdots & -\frac{1}{4}\frac{1}{5} \\ x-\frac{1}{2} & \cdots & x-\frac{1}{2} \end{pmatrix} \right.$$

$$\left. -\frac{1}{8}\frac{s-1}{2s-1} K\begin{pmatrix} -\frac{1}{16} & \cdots & -\frac{1}{16} \\ x-\frac{1}{2} & \cdots & x-\frac{1}{2} \end{pmatrix} K\begin{pmatrix} -\frac{(s-1)(s-3)}{(2s-3)(2s-5)} & \cdots & -\frac{1}{4}\frac{1}{5} \\ x-\frac{1}{2} & \cdots & x-\frac{1}{2} \end{pmatrix} \right]$$

which is the desired result.

Corollary 4.2.1: The variance function of the D_s-optimal design ξ^* is given by

$$d_s(x,\xi^*) = s - \frac{4^{m+r}[(2s-1)!!]^2}{s!(s-1)!} x(1-x)[p_r^{(\frac{1}{2},\frac{1}{2})}(x)p_{s-1}^{(1,1)}(x)$$

$$-\frac{1}{8}\frac{s-1}{2s-1} p_{r-1}^{(\frac{1}{2},\frac{1}{2})}(x)p_{s-2}^{(1,1)}$$

$$((2s-1)!! = (2s-1)(2s-3)\ldots 3.1).$$

If s = 1, then we have

$$d_1(x,\xi) = 1-4^{2m-1}x(1-x)[p_{m-1}^{(\frac{1}{2},\frac{1}{2})}(x)]^2.$$

Proof: It is obvious that

$$d_s(x,\xi^*) = s-cx(1-x)[p_r^{(\frac{1}{2},\frac{1}{2})}(x)p_{s-1}^{(1,1)}(x) - \frac{1}{8}\frac{s-1}{2s-1}p_{r-1}^{(\frac{1}{2},\frac{1}{2})}(x)p_{s-2}^{(1,1)}(x)]^2$$

where c is a constant to be determined. It is easily seen that $c = \frac{1}{\zeta_1\cdots\zeta_{2m}}$ and the result follows.

Corollary 4.2.2: If s = 1, then the support of the D_s-optimal design ξ^* is given by the zeros of $x(x-1)p_{m-1}^{(\frac{1}{2},\frac{1}{2})}(x)$. The corresponding weights on $0 = x_1 < \ldots < x_{m+1} = 1$ are 1 : 2 : 2 :...: 2 : 1.

Proof: The result follows from Corollary 4.2.1 and (4.2.2).

4.3 The D_s-Optimal Designs for w(x) = x, 1-x, x(1-x)

The results in this section are mainly from Studden (1981). We add improved expression for the support of the weighted D_s-optimal designs for the above weight functions. The extension of the results to general Jacobi weight functions seems extremely difficult.

Theorem 4.3.1: If w(x) = x, then $|\Sigma_s(\xi)|$ is maximized by

$$p_{2i} = \frac{1}{2} \qquad i = 1,2,\ldots,m$$

$$p_{2i+1} = \begin{cases} \frac{1}{2} & i = 0,1,\ldots,r \\ \frac{m-i+1}{2(m-i)+1} & i = r+1,\ldots,m-1 \end{cases}$$

$$p_{2m+1} = 1.$$

More explicitly, the weight attached to the point x_i is given by

$$\frac{2}{2m+2+\frac{\sin 2(r+1)\theta_i}{\sin \theta_i}} \tag{4.3.1}$$

where $2x_i - 1 = \cos \theta_i$, $0 \leq \theta_i \leq \pi$ and the x_i's are the zeros of

$$(x-1)[p_{r+1}^{(-\frac{1}{2},\frac{1}{2})}(x)p_{s-1}^{(1,1)}(x) - \frac{1}{8}\frac{s-1}{2s-1}p_r^{(-\frac{1}{2},\frac{1}{2})}(x)p_{s-2}^{(1,1)}(x)] = 0. \tag{4.3.2}$$

<u>Proof</u>: We only indicate a proof for (4.3.2).

The rest of the proof can be found in Studden (1981). The support of the D_s-optimal design is given by the zeros of

$$K\begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & \cdots & -\frac{1}{2} & -\frac{1}{2}\frac{s-1}{2s-1} & \cdots & -\frac{1}{2} \\ x & 1 \quad x & \cdots & x & 1 & \cdots & 1 \end{pmatrix}.$$

Factorizing out x-1, we get

$$(x-1)K\begin{pmatrix} -\frac{1}{4} & \cdots & -\frac{1}{4} & -\frac{1}{2}\frac{s-1}{2s-1} & -\frac{1}{2}\frac{s}{2s-1} & \cdots & -\frac{2}{3}\frac{1}{2} \\ x \quad 1 & \cdots & 1 & x & 1 & \cdots & x \quad 1 \end{pmatrix}.$$

Making an even contraction, we see that

$$(x-1)K\begin{pmatrix} -\frac{1}{16} & \cdots & -\frac{1}{16} & -\frac{1}{8}\frac{s-1}{2s-1} & -\frac{1}{4}\frac{s(s-2)}{(2s-1)(2s-3)} & \cdots & -\frac{1}{4}\frac{1}{5} \\ x-\frac{1}{4} \quad x-\frac{1}{2} & \cdots & x-\frac{1}{2} & x-\frac{1}{2} & & \cdots & x-\frac{1}{2} \end{pmatrix}.$$

Using Theorem 2.3.2 again, we obtain

$$(x-1)\left[K\begin{pmatrix} -\frac{1}{16} & \cdots & -\frac{1}{16} \\ x-\frac{1}{4} \quad x-\frac{1}{2} & \cdots & x-\frac{1}{2} \end{pmatrix} K\begin{pmatrix} -\frac{1}{4}\frac{s(s-2)}{(2s-1)(2s-3)} & \cdots & -\frac{1}{4}\frac{1}{5} \\ x-\frac{1}{2} & \cdots & x-\frac{1}{2} \end{pmatrix}\right.$$

$$- \frac{1}{8}\frac{s-1}{2s-1} K\begin{pmatrix} -\frac{1}{16} & \cdots & -\frac{1}{16} \\ x-\frac{1}{4} & x-\frac{1}{2}\cdots & x-\frac{1}{2}\end{pmatrix} K\begin{pmatrix} -\frac{1}{4}\frac{(s-1)(s-3)}{(2s-3)(2s-5)} & \cdots & -\frac{1}{4}\frac{1}{5} \\ x-\frac{1}{2} & \cdots & x-\frac{1}{2}\end{pmatrix}\Bigg]$$

which is the desired result.

Corollary 4.3.1:

$$d_s(x,\xi^*) = s- \frac{4^{m+r}[(2s-1)!!]^2}{(s+1)!s!}(1-x)[p_{r+1}^{(-\frac{1}{2},\frac{1}{2})}(x)p_{s-1}^{(1,1)}(x)$$

$$- \frac{1}{8}\frac{s-1}{2s-1} p_r^{(-\frac{1}{2},\frac{1}{2})}(x)p_{s-2}^{(1,1)}(x)]^2.$$

Proof: Similar to the proof for Corollary 4.2.1.

Corollary 4.3.2: If s = 1, then the support is given by the zeros of

$$(x-1)p_m^{(-\frac{1}{2},\frac{1}{2})}(x) = 0.$$

The corresponding weights on $x_1 < x_2 <\ldots< x_{m+1} = 1$ are 2: 2: ... : 2:1.

Proof: Follows from Corollary 4.3.1 and (4.3.1).

Similarly, we can prove the following results for $w(x) = 1-x$.

Theorem 4.3.2: If $w(x) = 1-x$, then $|\Sigma_s(\xi)|$ is maximized by

$$p_{2i} = \frac{1}{2} \qquad i = 1,2,\ldots,m$$

$$p_{2i+1} = \begin{cases} \frac{1}{2} & i = 0,1,\ldots,r \\ \frac{m-i}{2(m-i)+1} & i = r+1,\ldots,m-1 \end{cases}$$

$$p_{2m+1} = 0.$$

The support of ξ^*, the D_s-optimal design, are the zeros of

$$x(P_{r+1}^{(\frac{1}{2},-\frac{1}{2})}(x)P_{s-1}^{(1,1)}(x) - \frac{1}{8}\frac{s-1}{2s-1}P_r^{(\frac{1}{2},-\frac{1}{2})}(x)P_{s-2}^{(1,1)}(x).$$

The weight for x_i is given by

$$\xi^*(x_i) = \frac{2}{2m+2 + \frac{\sin 2(r+1)\theta_i}{\sin \theta_i}}$$

where $1-2x_i = \cos \theta_i$.

Proof: The proof is similar to that of Theorem 4.3.1.

Corollary 4.3.3:

$$d_s(x,\xi^*) = s - \frac{4^{m+r}[(2s-1)!!]^2}{(s-1)!s!} x[P_{r+1}^{(\frac{1}{2},-\frac{1}{2})}(x)P_{s-1}^{(1,1)}(x) - \frac{1}{8}\frac{s-1}{2s-1}P_r^{(\frac{1}{2},-\frac{1}{2})}(x)P_{s-2}^{(1,1)}(x)]^2.$$

Proof: The result follows by the same steps in proving Corollary 4.2.1.

Corollary 4.3.4: If $s = 1$, the support is given by the zeros of $x\,P_m^{(\frac{1}{2},-\frac{1}{2})}(x) = 0$. The corresponding weights on $0 = x_1 < x_2 < \ldots < x_{m+1}$ are 1: 2: 2: ... : 2.

Proof: Similar to the proof which was given for Corollary 4.3.2.

In the case $w(x) = x(1-x)$, we have the following results. The proofs are similar to those we gave for the previous cases and so are omitted.

Theorem 4.3.3: If $w(x) = x(1-x)$, then $|\Sigma_s(\xi)|$ is maximized by

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,m$$

$$p_{2i} = \begin{cases} \frac{1}{2} & i = 0,\ldots,r+1 \\ \frac{m-i+1}{2(m-i)+3} & i = r+2,\ldots,m \end{cases}$$

$$p_{2m+2} = 0.$$

The support of ξ^* is given by the zeros of

$$P_{r+2}^{(-\frac{1}{2},-\frac{1}{2})}(x)P_{s-1}^{(1,1)}(x) - \frac{1}{8}\frac{s-1}{2s-1} P_{r+1}^{(-\frac{1}{2},-\frac{1}{2})}(x)P_{s-2}^{(1,1)}(x) = 0.$$

The weights are given by

$$\xi(x_i) = \frac{2}{2m+3 - \dfrac{\sin(2r+3)\theta_i}{\sin\theta_i}} \qquad i = 1,\ldots,m+1$$

where $2x_i-1 = \cos\theta_i$.

Corollary 4.3.5:

$$d(x,\xi^*) = s - \frac{4^{m-r+1}[(2s-1)!!]^2}{s!(s-1)!} [P_{r+2}^{(-\frac{1}{2},-\frac{1}{2})}(x)P_{s-1}^{(1,1)}(x) - \frac{1}{8}\frac{s-1}{2s-1} P_{r+1}^{(-\frac{1}{2},-\frac{1}{2})}(x)P_{s-2}^{(1,1)}(x)]^2.$$

Corollary 4.3.6: If $s = 1$, the support of ξ^* is on the zeros of $P_{m+1}^{(-\frac{1}{2},-\frac{1}{2})}(x) = 0$. ξ^* assigns equal mass to each of these points.

It is difficult to extend the above method to the general Jacobi weight functions. We cannot even find the general solution for the simplest case $w(x) = x^2$, except for the particular instance $m = 1$ which is shown in the following example.

Example 4.3.1: Let $X = [0,1]$, $w(x) = x^2$ and $m = s = 1$. $|\Sigma_s(\xi)|$ can be written in the form

$$\frac{\begin{vmatrix}\mu_2 & \mu_3\\ \mu_3 & \mu_4\end{vmatrix}}{|\mu_2|} = \frac{\zeta_1^2(\zeta_2\zeta_3)(\zeta_1\zeta_3)}{\zeta_1(\zeta_1+\zeta_2)} .$$

After some tedious calculation, we can show that $p_1 = \frac{\sqrt{2}}{1+\sqrt{2}}$, $p_2 = \frac{1}{\sqrt{2}}$ and $p_3 = 1$. The D_s-optimal design is given by

$$\xi^*(\sqrt{2}-1) = \frac{1}{\sqrt{2}} \text{ and } \xi^*(1) = 1-\xi^*(\sqrt{2}-1).$$

4.4 D_s-Optimal Design for the s Highest Odd (Even) Coefficients

We assume throughout this section that the regression interval is $[-1,1]$. Let us first assume that the regression polynomial is of odd degree, i.e. m is odd. Suppose we want to estimate the s highest odd coefficients $\beta_{m-2s+2}, \beta_{m-2s+4},\ldots,\beta_m$. Before we give the main results, we first prove the following lemma.

Lemma 4.4.1: The D_s-optimal design ξ for the s highest odd (even) coefficients is symmetric. Here m may be odd or even.

Proof: Let $\xi'(x) = \xi(-x)$. It is easily seen that $|M(\xi)| = |M(\xi')|$ and $|M_{11}(\xi)| = |M_{11}(\xi')|$. Hence $|\Sigma_s(\xi)| = |\Sigma_s(\xi')|$. Using the fact that

$$\Sigma_s(\tfrac{1}{2}\xi + \tfrac{1}{2}\xi') \geq \tfrac{1}{2}\Sigma_s(\xi) + \tfrac{1}{2}\Sigma_s(\xi')$$

where equality holds if $M(\xi) = M(\xi')$ (see Kiefer's lecture notes), we see that

$$|\Sigma_s(\tfrac{1}{2}\,\xi + \tfrac{1}{2}\,\xi')| \geq |\tfrac{1}{2}\,\Sigma_s(\xi) + \tfrac{1}{2}\,\Sigma_s(\xi')|.$$

The convexity of -log det implies that

$$|\tfrac{1}{2}\,\Sigma_s(\xi) + \tfrac{1}{2}\,\Sigma_s(\xi')| \geq |\Sigma_s(\xi)|^{\frac{1}{2}}|\Sigma_s(\xi')|^{\frac{1}{2}} = |\Sigma_s(\xi)|.$$

So we have $M(\xi) = M(\xi')$ i.e. ξ is symmetric.

Lemma 4.4.2: Let $\{\mu_k\}_{k=0}^{4n+2}$ and $\{\mu_k'\}_{k=0}^{2n+1}$ be two sequences such that $\mu_{2k} = \mu_k'$ and $\mu_{2k+1} = 0$. Then

$$\begin{vmatrix} \mu_0 & 0 & \cdots & \mu_{2n} \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & & \vdots \\ \mu_{2n} & 0 & \cdots & \mu_{4n} \end{vmatrix} = \begin{vmatrix} \mu_0' & \mu_2' & \cdots & \mu_n' \\ \vdots & & & \vdots \\ \mu_n' & \mu_{n+1}' & \cdots & \mu_{2n}' \end{vmatrix} \begin{vmatrix} \mu_1' & \cdots & \mu_n' \\ \vdots & & \vdots \\ \mu_n' & \cdots & \mu_{2n-1}' \end{vmatrix}$$

$$\begin{vmatrix} \mu_0 & 0 & \cdots & \mu_{2n} & 0 \\ 0 & \mu_2 & \cdots & 0 & \mu_{2n+2} \\ \vdots & & & & \vdots \\ 0 & & \cdots & & \mu_{4n+2} \end{vmatrix} = \begin{vmatrix} \mu_0' & \cdots & \mu_n' \\ \vdots & & \\ \mu_n' & \cdots & \mu_{2n}' \end{vmatrix} \begin{vmatrix} \mu_1' & \cdots & \mu_{n+1}' \\ \vdots & & \vdots \\ \mu_{n+1}' & \cdots & \mu_{2n+1}' \end{vmatrix}.$$

Proof: Use the Laplacian expansion of the determinant.

Theorem 4.4.1: The D_s-optimal design for the s highest odd coefficients when m is odd is given by

$$p_{odd} = \frac{1}{2}$$

$$p_{2i} = \begin{cases} \frac{1}{2} & i \text{ is either even or } i = 1,3,\ldots,m-2s \\ \frac{m-i+2}{2(m-i+1)} & i = m-2s+2,\ldots,m-2 \end{cases}$$

$$p_{2m} = 1.$$

The support of the D_s-optimal design is given by the zeros of

$$(x^2-1)\left[P^{(-\frac{1}{2},\frac{1}{2})}_{\frac{m+1}{2}-s}(x^2)P^{(1,1)}_{s-1}(x^2) - \frac{1}{8}\,\frac{s-1}{2s-1}\,P^{(-\frac{1}{2},\frac{1}{2})}_{\frac{m+1}{2}-s}(x^2)P^{(1,1)}_{s-2}(x^2)\right].$$

The weight at x_i is given by

$$\frac{1}{m+1+\dfrac{\sin(m-2s+1)\theta_i}{\sin\theta_i}}$$

where $2x_i^2-1 = \cos\theta_i$, $0 \le \theta_i \le \pi$.

Proof: By Lemma 1 and Lemma 2, we see that

$$|\Sigma_s(\xi)| = \frac{\begin{vmatrix} \mu'_1 & \mu'_2 & \cdots & \mu'_{\frac{m+1}{2}} \\ \mu'_2 & \mu'_3 & \cdots & \mu'_{\frac{m+2}{2}} \\ \vdots & & & \vdots \\ \mu'_{\frac{m+1}{2}} & \mu'_{\frac{m+3}{2}} & \cdots & \mu'_m \end{vmatrix}}{\begin{vmatrix} \mu'_1 & \cdots & \mu'_{\frac{m+1}{2}-s} \\ \vdots & & \vdots \\ \mu'_{\frac{m+1}{2}-s} & \cdots & \mu'_{m+2s} \end{vmatrix}} \tag{4.4.1}$$

where μ'_k is the k-th moment of ξ', and ξ' is a design on [0,1] which is related to a design ξ on [-1,1] by the following relation

$$\xi'(x^2) = \xi(x) + \xi(-x) = 2\xi(x). \tag{4.4.2}$$

From (4.4.1) we see that the problem is reduced to finding the D_s-optimal design ξ' on [0,1] with $w(x) = x$. The solution to the original problem ξ can be recovered by (4.4.2). The result now follows from Theorem 4.3.1.

Similarly we can solve the problem of finding the D_s-optimal design for the s highest even coefficients when m is even. In this case the problem can be reduced to finding the D_s-optimal design for the s highest coefficients with the degree of the regression polynomial given by $\frac{m}{2}$ and the regression region equal to [0,1]. The result can then be found by using Theorem 4.2.1. So we have proved

Theorem 4.4.2: The D_s-optimal design ξ for the s highest even coefficients when m is even is given by

$$p_{odd} = \frac{1}{2}$$

$$p_{2i} = \begin{cases} \frac{1}{2} & i \text{ odd or } i = 2,4,\ldots,m-2s \\ \frac{m-i+2}{2(m-i+1)} & i = m-2s+2,\ldots,m-2. \end{cases}$$

The support of ξ is given by the zeros of

$$x(x^2-1)[P_r^{(\frac{1}{2},\frac{1}{2})}(x^2)P_{s-1}^{(1,1)}(x^2) - \frac{1}{8}\frac{s-1}{2s-1}P_{r-1}^{(\frac{1}{2},\frac{1}{2})}(x^2)P_{s-2}^{(1,1)}(x^2)]$$

The weight at x is equal to

$$\begin{cases} \frac{1}{m+1+U_{2r}(x)} & \text{if } x = 0 \\ \frac{x}{m+1+U_{2r}(x)} & \text{if } x \neq 0. \end{cases}$$

Remark 4.4.1: It should be noted that the problem for finding the D_s-optimal design for the s highest odd coefficients when m is even can be reduced to the case where m is odd by ignoring the last coefficient. When m is odd the similar remark applies when finding the D_s-optimal design for the s highest even coefficients.

It is natural to insquire about the D_s-optimal design for an arbitrary subset of the β-vector. It seems that the problem is quite complicated except in some simple settings as the following examples show.

Example 4.4.1: Let $x = [-1,1]$ and $E(y) = \beta_0 + \beta_1 x$. It is easily seen that the D_s-optimal design for estimating β_1 is identical to the D-optimal design. Any design that is symmetric about zero is good for estimating β_0.

Example 4.4.2: Let $x = [-1,1]$ and $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$. The D_s-optimal design for β_1 and β_2 is identical with the D-optimal design. The D_s-optimal designs for $\{\beta_0,\beta_2\}$, $\{\beta_4,\beta_0\}$ and $\{\beta_2\}$ are identical. The D_s-optimal design for β_1 is identical with the D_s-optimal design for β_1 in Example 4.3.1. The D_s-optimal design for β_0 is just the design that puts all mass at $x = 0$.

Example 4.4.3: Let $x = [-1,1]$ and $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$. This is the lowest degree model in which not every combination of the β_i's has an easily determined D_s-optimal design. For example, it is hard to compute the D_s-optimal design for $\{\beta_0,\beta_1\}$.

4.5 D_s-Optimal Designs for $w(x) = (1-x^2)x^2$ and $w(x) = x^2$

In this section we assume that $x = [-1,1]$. Using the same argument as in Section 4.4, we see that the D_s-optimal design for the s highest odd (even) coefficients is symmetric when $w(x) = (1-x^2)x^2$ or $w(x) = x^2$. Using the same technique we used in the last section, we have the following results.

<u>Theorem 4.5.1</u>: Let $x = [-1,1]$, $w(x) = (1-x^2)x^2$ and $f^T(x) = \sqrt{w(x)}\,(1,x,\ldots,x^m)$ where m is even. The D_s-optimal design for the s highest even coefficients is given by

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,m+1$$

$$p_{2i} = \begin{cases} \frac{1}{2} & i \text{ is odd or } i = 2,4,\ldots,m-2s+2 \\ \frac{m-i+2}{2(m-i+1)} & i = m-2s+4,\ldots,m \end{cases}$$

$$p_{2m+4} = 0.$$

<u>Proof</u>: By Lemma 4.4.1 and Lemma 4.4.2, we find that

$$|\Sigma_s(\xi)| = \frac{\left|\int_{-1}^{1} (1-x^2)x^{2+i+j}d\xi\right|_{i,j=0}^{m}}{\left|\int_{-1}^{1} (1-x^2)x^{2+i+j}d\xi\right|_{i,j=0}^{m-2s}} . \qquad (4.5.1)$$

By making the transformations $t = x^2$ and $\xi'(t) = \xi(x) + \xi(-x)$, we can rewrite (4.5.1) as

$$\frac{\left|\int_{0}^{1} t(1-t)t^{i+j}d\xi'\right|_{i,j=0}^{\frac{m}{2}}}{\left|\int_{0}^{1} t(1-t)t^{i+j}d\xi'\right|_{i,j=0}^{\frac{m}{2}-s}} . \qquad (4.5.2)$$

By Theorem 2.2.2, (4.5.2) can be written as

$$\frac{\prod_{i=1}^{\frac{m}{2}+1} (\gamma_{2i-1}\gamma_{2i})^{\frac{m}{2}+2-i}}{\prod_{i=1}^{\frac{m}{2}-s+1} (\gamma_{2i-1}\gamma_{2i})^{\frac{m}{2}-s+2-i}}$$

which can be simplified as

$$\prod_{i=1}^{\frac{m}{2}-s+1} (\gamma_{2i-1}\gamma_{2i})^{s} \prod_{i=\frac{m}{2}-s+2}^{\frac{m}{2}+1} (\gamma_{2i-1}\gamma_{2i})^{\frac{m}{2}+2-i} .$$

It is maximized if

$$p_i = \frac{1}{2} \qquad i \text{ is odd or } i = 2,4,\ldots,m-2s+2$$

$$p_{2i} = \frac{m-2i+2}{2(m-2i+1)} \qquad i = \frac{m}{2}-s+2,\ldots,\frac{m}{2}$$

$$p_{m+2} = 0.$$

Converting it back to [-1,1], we obtain the result.

Immediately we have the following corollary.

Corollary 4.5.1: If $s = 1$, we have

$$p_i = \frac{1}{2}, \qquad i = 1,2,\ldots,2m+3$$

$$p_{2m+4} = 0.$$

The support of the design is given by the zeros of $T_{m+2}(x) = 0$ where $T_{m+2}(x)$ is the (m+2)th Tchebycheff polynomial. The weight attached to each point is $\frac{1}{m+2}$.

The case m is odd can be reduced to that of finding the weighted D_s-optimal design with weight function $w(x) = x^2$.

Example 4.5.1: The case $m = 3$, $s = 1$ and $w(x) = (1-x^2)x^2$ can be reduced to the D_s-optimal design with $w(x) = x^2$ with $m = s = 1$ and $X = [0,1]$. The solution has been given in Example 4.3.1.

The same reasoning leads us to the following results.

Theorem 4.5.2: Let $x = [-1,1]$, $w(x) = x^2$ and $f^T(x) = \sqrt{w(x)}\,(1,x,\ldots,x^m)$ where m is even. The D_s-optimal design for the s highest even coefficients is given by

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,m$$

$$p_{2i} = \begin{cases} \frac{1}{2} & i \text{ is even or } i = 1,3,\ldots,m-2s+1 \\ \frac{m-i+3}{2(m-i+2)} & i = m-2s+3,\ldots,m-1 \end{cases}$$

$$p_{2m+2} = 1.$$

Corollary 4.5.2: If $s = 1$ in the above theorem, we have

$$p_i = \frac{1}{2} \qquad i = 1,2,\ldots,2m+1 \quad \text{and}$$

$$p_{2m+2} = 1.$$

The support of the D_s-optimal design is given by the zeros of

$$(x^2-1)U_m(x) = 0$$

where $U_m(x)$ is the m-th Tchebycheff polynomial of the second kind.

4.6 D_s-Optimal Design for Trigonometric Regression

Consider the trigonometric regression model

$$E(y) = \alpha_0 + \sum_{n=1}^{m} (\alpha_n \cos n\theta + \beta_n \sin n\theta).$$

The problem of interest is to find a D_s-optimal design for the s highest pairs of coefficients. We have the following theorem.

Theorem 4.6.1: A D_s-optimal design for the s highest pairs of coefficients is characterized by the condition

$$a_0 = a_1 = \ldots = a_{2m-1} = 0$$

where a_i is a parameter of measure on the circle which is defined in Chapter II Section 6.

Proof: It is easy to see that

$$|\Sigma_s(\xi)| = \frac{\Delta_{2m}}{\Delta_{2m-2s}} \times \text{const.}$$

where Δ_{2k} denotes the Toeplitz determinant of order k+1. By (2.6.8) we see that $|\Sigma_s(\xi)|$ can be expressed as

$$\frac{\prod_{i=0}^{2m-1} (1-|a_i|^2)^{2m-i}}{\prod_{i=0}^{2m-2s-1} (1-|a_i|^2)^{2m-2s-i}}$$

$$= \prod_{i=0}^{2m-2s-1} (1-|a_i|^2)^{2s} \prod_{i=2m-2s}^{2m-1} (1-|a_i|^2)^{2m-i}.$$

The preceding quantity is maximized iff

$$a_0 = a_1 = \ldots = a_{2m-1} = 0.$$

Remark 4.6.1: The D_s-optimal design coincides with the D-optimal design for the trigonometric regression.

4.7 Rotatable D_s-Optimal Design

The rotatable D_s-optimal design is difficult to find in general except in some simple cases that we will indicate below.

For the 1st degree regression (d=1) on the m-ball it is obvious the rotatable D_s-optimal design (s=1) is the same as the rotable D-optimal design. (For the notation see Chapter III, Section 6.)

Example 4.7.1: Let d = 2; i.e. 2nd degree regression on the m-ball. $|\Sigma_s(\xi)|$ can be expressed as

$$\frac{p_1^{1+m+\frac{(m-1)(m+2)}{2}} q_1 p_2 (p_1+q_1p_2)^{\frac{(m-1)(m+1)}{2}}}{p_1^m}.$$

The above quantity is maximized by

$$p_1 = \frac{m(m+1)}{m^2+m+2}$$

$$p_2 = 1.$$

This means that we have to put mass $\frac{3}{4}$ on the outer boundary and the rest on the center.

4.8 D_s-Optimal Design (s=1) for $w(x) = \frac{1}{w_n(x)}$ Where $w_n(x)$ Is a Positive Polynomial on [-1,1].

Let c be a (m+1)-vector and consider estimating $c^T\beta$ where β is the vector of coefficients of the regression polynomial. A vector c is said to be estimable with respect to the design iff c belongs to the space generated by the set of vectors $\{(1,x,\ldots,x^m)|x \in s(\xi)$ (spectrum of ξ)$\}$. If c is estimable with respect to ξ, let

$$d(c,\xi) = \sup \frac{(c^Td)^2}{d^TM(\xi)d}$$

where the sup is taken over the set of vectors d such that the denominator is nonzero. If c is not estimable with respect to ξ, we define $d(c,\xi) = \infty$. It is known that the variance of the best linear unbiased estimate of $c^T\beta$ is proportional to $d(c,\xi)$.

Definition 4.8.1: A design ξ is said to be c-optimal if ξ minimizes $d(c,\xi)$.

Theorem 4.8.1: (Elfving's Theorem). A design $\xi^* = \{x_v, p_v\}$ minimizes $d(c,\xi)$ iff there exists $\varepsilon_v = \pm 1$ such that

$$\sum_{v=1}^{k} \epsilon_v p_v f(x_v) = \lambda c \in \partial \mathcal{R} \text{ (= boundary of } \mathcal{R})$$

where the integer k may always be taken to be at most m+2 and at most m+1 if c is a boundary point of $\mathcal{R}$. Here $\mathcal{R}$ is the convex hull of $\{f(x) \mid x \in X\} \cup \{-f(x) \mid x \in X\}$ and X is the regression region. Moreover $d(c,\xi) = \frac{1}{\lambda^2}$.

Proof: See Karlin and Studden (1966b).

Lemma 4.8.1: $\sum_v \epsilon_v p_v f(x_v) = \lambda c \in 2\mathcal{R}$ iff there exists a nontrivial polynomial $u(x) = \sum_i a_i f_i(x)$ such that $|u(x)| \leq \lambda$ for all u, $\varepsilon_v\, u(x_v) = \lambda$ and $c^T a = 1$.

Proof: See Studden (1968).

Now, observe that

$$\min_{\xi} d(c,\xi) = \min_{\xi} \sup_{d} \frac{(d^T c)^2}{d^T M(\xi) d}$$

$$= \min_{\xi} \sup_{d} \frac{(d^T c)^2}{\int_X (d^T f(x))^2 d\xi(x)}$$

$$\geq \sup_{d} \min_{\xi} \frac{(d^T c)^2}{\int_X (d^T f(x))^2 d\xi(x)}.$$

We can restrict ourselves to those d satisfying $d^T c = 1$, hence the above expression can be rewritten as

$$\frac{1}{\inf_{\xi} \max_{x} \int_X (d^T f(x))^2 d\xi(x)}$$

$$\geq \frac{1}{\inf_{d} \max_{x} |d^T f(x)|^2}$$

$$= \frac{1}{\lambda^2}.$$

Since $\min_{\xi} d(c,\xi) = \frac{1}{\lambda^2}$, equalities hold. It is clear now that the support of the c-optimal design consists of points that satisfy the equation

$$|d^T f(x)| = \lambda.$$

Notice that the D_s-optimal (s=1) design for the highest coefficient must coincide with the c-optimal design when c = (0,0,...,1). In this case $d^T f(x)$ is the "polynomial" of least deviation from zero. In Theorem 2.8.3, the "polynomials" of least deviation from zero are given for

$$f(x) = \frac{1}{\sqrt{w_n(x)}}(1,\ldots,x^m),\ \sqrt{\frac{1-x^2}{w_n(x)}}(1,\ldots,x^m),$$

$$\sqrt{\frac{1-x}{w_n(x)}}(1,\ldots,x^m) \text{ and } \sqrt{\frac{1+x}{w_n(x)}}(1,\ldots,x^m).$$

So we have the following theorem.

<u>Theorem 4.8.2</u>: (1) Let $w(x) = \frac{1}{w_n(x)}$. The D_s-optimal design (s=1) for estimating the highest coefficient has its support resting on the zeros of $(x^2-1)U_{m-1}(x, \frac{1}{w_n(x)})$ where $U_k(x, \frac{1}{w_n(x)})$ is the k-th orthogonal

polynomial with respect to $\frac{\sqrt{1-x^2}}{\pi w_n(x)}$ and $2m \geq n$. (2) Let $w(x) = \frac{1-x^2}{w_n(x)}$. The D_s-optimal design (s=1) for estimating the highest coefficient has its support resting on the zeros of $T_{m+1}(x, \frac{1}{w_n(x)})$ where $T_k(x, \frac{1}{w_n(x)})$ is the k-th orthogonal polynomial with respect to $\frac{1}{\pi w_n(x)\sqrt{1-x^2}}$ and $2m \geq n$.

(3) Let $w(x) = \frac{1-x}{w_n(x)}$. The D_s-optimal design (s = 1) for estimating the highest coefficient has its support resting on the zeros of $(1+x)W_m(x, \frac{1}{w_n(x)})$ where $W_k(x, \frac{1}{w_n(x)})$ is the k-th orthogonal polynomial with respect to $\frac{1}{\pi w_n(x)} \sqrt{\frac{1-x}{1+x}}$ and $2m \geq n$.

(4) Let $w(x) = \frac{1+x}{w_n(x)}$. The D_s-optimal design (s=1) for estimating the highest coefficient has its support resting on the zeros of $(x-1)V_m(x, \frac{1}{w_n(x)})$ where $V_k(x, \frac{1}{w_n(x)})$ is the k-th orthogonal polynomial with respect to $\frac{1}{\pi w_n(x)} \sqrt{\frac{1+x}{1-x}}$ and $2m \geq n$.

CHAPTER V

I_σ-OPTIMAL DESIGN AND OPTIMAL EXTRAPOLATION DESIGN

5.1 I_σ-Optimal Design

Recall that the variance function of a design ξ at a point x is defined as

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

Notice that $d(x,\xi)$ is a function in x in general. Sometimes the accuracy required at each point is different and some weighted "sum" of the variances at different x's may serve as a optimality criterion. Thus it seems reasonable to minimize

$$v(\xi,\sigma) = \int d(x,\xi)d\sigma \tag{5.1.1}$$

over all designs ξ for a fixed probability measure σ. A design ξ^* which minimizes $v(\xi,\sigma)$ will be called an I_σ-optimal design. Some results have been obtained about the I_σ-optimal design by Studden (1971), (1977) and Federov (1972). Since the I_σ-optimal design is quite difficult to obtain in general, Studden (1976) suggested using an approximate design. Some asymptotic results have been obtained in the same paper. In this section we will find I_σ-optimal designs for some special σ and try to express $v(\xi,\sigma)$ in terms of canonical moments. Some asymptotic results will be given in Chapter VII. Let $g(x) = Tf(x)$ where T is a nonsingular matrix of size m+1. It is clear that

$$\int f(x)^T M^{-1}(\xi) f(x) d\sigma(x) = \int g(x)^T M^{-1}(\xi) g(x) d\sigma(x).$$

Thus we have

Theorem 5.1.1: The I_σ-optimal design for f(x) coincides with the I_σ-optimal design for g(x).

Corollary 5.1.1: If $f(x)^T = \sqrt{w(x)}\,(1,x,\ldots,x^m)$, then the I_σ-optimal design remains the same under a linear transformation on the regression interval (assuming w(x) undergoes the same transformation.)

If we write trA for the trace of the matrix A, then (5.1.1) can be written as

$$v(\xi,\sigma) = \mathrm{tr}M^{-1}(\xi)M(\sigma).$$

The next theorem gives a necessary and sufficient condition for the I_σ-optimal design.

Theorem 5.1.2: The design ξ is I_σ-optimal iff

$$f^T(x)M^{-1}(\xi)M(\sigma)M^{-1}(\xi)f(x) \leq \mathrm{tr}M^{-1}(\xi)M(\sigma). \qquad (5.1.2)$$

Proof: See Fedorov (1972).

Immediately, we have the following theorem.

Theorem 5.1.3: Suppose σ has arbitrary weights $q_0,\ldots,q_m$, $\sum_{i=0}^{m} q_i = 1$, on the support of the weighted D-optimal design. Then the I_σ-optimal design ξ has the same support as σ with weights proportional to $\sqrt{q_i}$, $i = 0,1,\ldots,m$.

Proof: Notice that we have assumed that the weighted D-optimal design has its support resting on m+1 points. Let $f_i(x) = \ell_i(x)$ where $\ell_i(x)$ denotes the Lagrange function corresponding to the support of the weighted D-optimal design $\{x_0,\ldots,x_m\}$. It can be easily seen that (5.1.2) becomes

$$w(x) \sum_{i=0}^{m} \frac{\ell_i^2(x)}{w(x_i)} \leq 1.$$

The above equation holds by Theorem 3.1.3; the proof is completed.

Corollary 5.1.3: Let $X = [0,1]$. The above result holds for the following weight functions

(1) $w(x) \equiv 1$

(2) $w(x) = x^{\alpha+1}(1-x)^{\beta+1}$, $\alpha > -1$, $\beta > -1$

(3) $w(x) = x^{\alpha+1}$ $\alpha > -1$

(4) $w(x) = (1-x)^{\beta+1}$, $\beta > -1$

(5) $w(x) = [x(1-x)]^{\alpha+1}|x - \frac{1}{2}|^{\gamma}$, $\alpha \geq -1$, $\gamma > 0$ (m is odd in this case).

Theorem 5.1.4: Let $f^T(x) = (1, 2\cos x, 2\sin x,\ldots,2\cos mx, 2\sin mx)$ and $X = [0,2\pi]$. If σ is characterized by $a_i = 0$, $i = 0,1,\ldots,2m-1$, then any design characterized by the same condition is I_σ-optimal.

Proof: Use (5.1.2).

The following examples show that the integrated variance $v(\xi,\sigma)$ can be expressed in terms of canonical moments of ξ and σ.

Example 5.1.1: Let m = 1. The integrated variance $v_1(\xi,\sigma)$ can be written

$$v_1(\xi,\sigma) = \int_0^1 d_1(x,\xi)d\sigma$$

where $d_1(x,\xi)$ is the variance function of the design ξ for simple linear regression. It is known that $d_1(x,\xi)$, can be expressed as sum of squares of orthonormal polynomials.

$$d_1(x,\xi) = 1 + \frac{1}{p_1q_1p_2} P_1^2(x,d\xi)$$

where $\{P_k(x,d\xi)\}$ are monic orthogonal polynomials. Since $P_1(x,d\xi) = x-p_1$, $v_1(\xi,\sigma)$ can be written

$$\int_0^1 [1+ \frac{1}{p_1q_1p_2} (x-p_1)^2]d\sigma(x).$$

Let α_i denote the i-th canonical moments of σ and let $\beta_i = 1-\alpha_i$. Then we have

$$v_1(\xi,\sigma) = \int_0^1 1+ \frac{1}{p_1q_1p_2} (x-\alpha_1+\alpha_1-p_1)^2 d\sigma$$

$$= 1 + \frac{1}{p_1q_1p_2} [(p_1-\alpha_1)^2 + \alpha_1\beta_1\alpha_2].$$

To minimize the preceding expression, we need to take $p_2 = 1$ and $p_1 = \dfrac{\sqrt{\alpha_1(\alpha_1+\beta_1\alpha_2)}}{\sqrt{\alpha_1(\alpha_1+\beta_1\alpha_2)}+\sqrt{\beta_1(\beta_1+\alpha_1\alpha_2)}}$.

Example 5.1.2: Let m = 2. $d_2(x,\xi)$ can be written

$$d_2(x,\xi) = 1 + \frac{1}{p_1q_1p_2} P_1^2(x,d\xi) + \frac{1}{p_1q_1p_2q_2p_3q_3p_4} P_2^2(x,\xi).$$

As in Example 5.1.1, $P_1(x,d\xi)$ can be written

$$P_1(x,d\xi) = P_1(x,d\sigma) + (\alpha_1-p_1).$$

Similarly we can write $P_2(x,d\xi)$ as

$$P_2(x,d\xi) = P_2(x,d\sigma) + (\sum_{i=1}^{3} \zeta_i' - \sum_{i=1}^{3}\zeta_i)P_1(x,d\sigma)$$
$$+ \zeta_3(\zeta_1-\zeta_1')-\zeta_1'(\zeta_1'+\zeta_2'-\zeta_1-\zeta_2).$$

Here $\{P_k(x,d\sigma)\}$ are (monic) orthogonal with respect to $d\sigma$ and $\zeta_i' = \beta_{i-1}\alpha_i$ for $i \geq 2$ and $\zeta_1' = \alpha_1$. It follows that

$$v_2(\xi,\sigma) = 1 + \frac{1}{\zeta_1\zeta_2}[(\zeta_1-\zeta_1')^2 + \zeta_1'\zeta_2'] + \frac{1}{\zeta_1\zeta_2\zeta_3\zeta_4}\{\zeta_1'\zeta_2'\zeta_3'\zeta_4' +$$
$$(\zeta_1'+\zeta_2'+\zeta_3'-\zeta_1-\zeta_2-\zeta_3)^2\zeta_1'\zeta_2'+[\zeta_3(\zeta_1-\zeta_1')-\zeta_1'(\zeta_1'+\zeta_2'-\zeta_1-\zeta_2)]^2\}.$$

It should be remarked that for given ξ, $v_k(\xi,\sigma)$ is a nondecreasing function in k. If we let $p_i = \alpha_i$ for $i \leq 2k-1$, $p_{2k} = 1$, then

$$v_k(\xi,\sigma) \leq k + \alpha_{2k}.$$

5.2 Weighted Extrapolation Design

Let $w(x)$ be a continuous nonnegative weight function on $[a,b] \subseteq \mathbb{R}$ and let $f(x) = \sqrt{w(x)}\,(1,x,\ldots,x^m)$. We want to find a design ξ such that

$$d(z_0,\xi) = f(z_0)^T M^{-1}(\xi) f(z_0)$$

is minimized. The resulting design is called the weighted optimal extrapolation design. This is a special type of c-optimal design with $c^T = (1,z_0,\ldots,z_0^m)$ where $z_0 \notin [a,b]$. According to the discussion in Chapter 4, we see that it is equivalent to finding a polynomial $a^T f(x) = u(x)$ such that $\max|u(x)|$ is minimized subject to $a^T c = 1$. The support of the design is the set of all x such that $\max|u(x)|$ is attained. Once the support of the optimal extrapolation design, say $x_0,\ldots,x_m$ is given, the weight attached to each point can be calculated by the following method. Let

$$F\binom{0,1,\ldots,m}{x_0 x_1,\ldots,x_m} = \begin{vmatrix} f_0(x_0) & f_0(x_1) & \cdots & f_0(x_m) \\ f_1(x_0) & f_1(x_1) & \cdots & f_1(x_m) \\ \vdots & & & \vdots \\ f_m(x_0) & f_m(x_1) & \cdots & f_m(x_m) \end{vmatrix},$$

and

$$L_v(z_0) = \frac{F\binom{0,\ldots,\ v-1,\ v,\ v+1,\ldots,m}{x_0,\ldots,x_{v-1},z_0,x_{v+1},\ldots,x_m}}{F\binom{0,\ldots,m}{x_0,\ldots,x_m}},$$

where $f_i(x) = \sqrt{w(x)}\, x^i$, $i = 0,1,\ldots,m$, then the weight at x_i is given by

$$\xi_i = \frac{|L_i(z_0)|}{\sum_{v=0}^{m} |L_v(z_0)|}. \tag{5.2.1}$$

It is easy to observe that the D_s-optimal design when $s = 1$ coincides with the c-optimal design when $c = (0,0,\ldots,0,1)$. On the other hand, according to the discussion in Chapter 4, the support of the optimal extrapolation design are those x such that

$$\min_a \max_x |a^T f(x)| \text{ is attained}$$

where $c^T a = 1$ with $c^T = (1,z_0,\ldots,z_0^m)$, $z_0 \notin [a,b]$. Compare to the similar formulation for the case $c^T = (0,0,\ldots,1)$, we conclude that the optimal extrapolation design has the same support as the D_s-optimal design for $s = 1$.

In case we are interested in finding the weighted optimal extrapolation design for $\beta^T g(z_0)$ where $z_0 \notin [a,b]$ or $z_0 \in [a,b]$ with

$w(z_0) = 0$, and $g^T(z_0) = (1, z_0, \ldots, z_0^m)$. The criterion function is

$$d(z_0,\xi) = g^T(z_0, M^{-1}(\xi) g(z_0)$$
$$= \sup_d \frac{(g(z_0),d)}{(d,M(\xi)d)}.$$

Following the same procedure in 4.8 we see that it is equivalent to finding a polynomial that assume one at the point z_0 and has minimum maximum (weighted) deviation from 0. So we see that the support of the optimal extrapolation design is the same as we have found above. But the weights are given by

$$\xi_i = \frac{|L_i(z_0)| \frac{1}{\sqrt{w(x_i)}}}{\sum_{i=0}^{m} |L_i(z_0)| \frac{1}{\sqrt{w(x_i)}}} \qquad i = 0,1,\ldots,m$$

where

$$L_v(z_0) = \frac{F\binom{0,\ldots,v-1,\ v,\ v+1,\ldots,m}{x_0,\ldots,x_{v-1},z_0,x_{v+1},\ldots,x_m}}{F\binom{0,\ldots,m}{x_0,\ldots,x_m}}$$

$$F\binom{0,1,\ldots,m}{x_0,x_1,\ldots,x_m} = \begin{vmatrix} g_0(x_0) & g_0(x_1) & \cdots & g_0(x_m) \\ g_1(x_0) & g_1(x_1) & \cdots & g_1(x_m) \\ \vdots & & & \vdots \\ g_m(x_0) & g_m(x_1) & \cdots & g_m(x_m) \end{vmatrix}.$$

Theorem 5.2.1: (i) Let $X = [0,1]$ and $w(x) = x$. The support of the optimal extrapolation design is given by the zeros of $(x-1)P_m^{(-\frac{1}{2},\frac{1}{2})}(x) = 0$.

(ii) Let $X = [0,1]$ and $w(x) = 1-x$. The support of the optimal extrapolation design is given by the zeros of $x\, P_m^{(\frac{1}{2},-\frac{1}{2})}(x) = 0$.

(iii) Let $X = [0,1]$ and $w(x) = x(1-x)$. The support of the optimal extrapolation design is given by the zeros of $P_{m+1}^{(-\frac{1}{2},-\frac{1}{2})}(x) = 0$.

(iv) Let $X = [-1,1]$ and $w(x) = (1-x^2)|x|^2$ and m is even. The support of the optimal extrapolation design is given by the zeros of $T_{m+2}(x) = 0$ where $T_k(x)$ is the k-th Tchebycheff polynomial.

(v) Let $X =[-1,1]$ and $w(x) = x^2$ and m is even. The support of the optimal extrapolation design is given by the zeros of $(1-x^2)U_m(x) = 0$.

(vi) Let $X = [-1,1]$ and $w(x) = \frac{1}{w_n(x)}$. The support of the extrapolation design are the zeros of $(x^2-1)U_{m-1}(x, \frac{1}{w_n(x)}) = 0$.

(vii) Let $X = [-1,1]$ and $w(x) = \frac{1-x^2}{w_n(x)}$. The support of the optimal extrapolation design is given by the zeros of $T_{m+1}(x, \frac{1}{w_n(x)}) = 0$.

(viii) Let $X = [-1,1]$ and $w(x) = \frac{1-x}{w_n(x)}$. The support of the optimal extrapolation design is given by the zeros of $(1+x)W_m(x, \frac{1}{w_n(x)}) = 0$.

(ix) Let $X = [-1,1]$ and $w(x) = \frac{1+x}{w_n(x)}$. The support of the optimal extrapolation design is given by the zeros of $(1-x)V_m(x, \frac{1}{w_n(x)}) = 0$.

The following example is taken from Fedorov (1972).

Example 5.2.1: Let $E(y|x) = \beta_0 + \beta_1 x$ and $w(x) = x^2$, $0 \leq x \leq 1$. It is required to find the design minimizing $d(0,\xi)$. The regression problem formulated is very often met in the investigation of the scattering of elementary particles. The support of the optimal extrapolation design was found in Example 4.3.1, supp(ξ) = $\{\sqrt{2}-1,1\}$. According to (5.2.1), the weights at these two points are given by

$$\xi(\sqrt{2}-1) = \frac{2+\sqrt{2}}{4}$$

$$\xi(1) = \frac{2-\sqrt{2}}{4}.$$

5.3 The Stieltjes Transform and Canonical Moments of the Optimal Extrapolation Design (w(x) = 1).

Let $X = [-1,1]$ and $c^T = f(z_0) = (1,z_0,\ldots,z_0^m)$ where $|z_0| > 1$. Hoel and Levine (1964) showed that the optimal extrapolation at z_0 is given by $\xi(x_i) = \dfrac{|L_i(z_0)|}{\sum_{j=0}^{m} |L_j(z_0)|}$ where $x_v = \cos \frac{v}{m}\pi$, $v = 0,1,\ldots,m$ and $L_i(x)$, $i = 0,1,\ldots,m$ is the i-th Lagrange interpolation polynomial. In this section we will show how to find the Stieltjes transform and canonical moments of the optimal extrapolation design. With this formula we can prove some limt theorems.

Theorem 5.3.1: Let ξ denote the optimal extrapolation design when $w(x) = 1$. Then we have

$$\int_{-1}^{1} \frac{d\xi(x)}{z-x} = \cfrac{1}{z-\frac{T_{m-1}}{T_m}} - \cfrac{\frac{T_{m-1}}{T_m}\left(z_0-\frac{T_{m-1}}{T_m}\right)}{z-\frac{1}{2}\frac{T_{m-2}}{T_{m-1}}+\frac{T_{m-1}}{T_m}} - \cfrac{\frac{1}{4}\frac{T_{m-2}T_m}{T_{m-1}^2}}{z-\frac{1}{2}\left(\frac{T_{m-3}}{T_{m-2}}-\frac{T_{m-2}}{T_{m-1}}\right)}$$

$$- \cfrac{\frac{1}{4}\frac{T_{m-3}T_{m-1}}{T_{m-2}^2}}{z-\frac{1}{2}\left(\frac{T_{m-4}}{T_{m-3}}-\frac{T_{m-3}}{T_{m-2}}\right)} - \cdots - \cfrac{\frac{1}{4}\frac{T_2T_4}{T_3^2}}{z-\frac{1}{2}\left(\frac{T_1}{T_2}-\frac{T_2}{T_3}\right)} - \cfrac{\frac{1}{4}\frac{T_1T_3}{T_2^2}}{z-\frac{1}{2}\left(\frac{T_0}{T_1}-\frac{T_1}{T_2}\right)}$$

$$- \cfrac{\frac{1}{4}\frac{T_2}{T_1}}{z+\frac{1}{2T_1}} . \tag{5.3.1}$$

Here T_k denote the k-th Tchebycheff polynomial evaluated at z_0.

Before we proceed to prove Theorem 5.3.1, we first give the following results.

Lemma 5.3.1: $\sum_{v=0}^{m} |L_v(z_0)| = |T_m(z_0)|$.

Proof: See Karlin and Studden (1966a).

Lemma 5.3.2: (1) $\frac{d}{dx}(x^2-1)U_{m-1}(x)\Big|_{x=x_v} = U'_{m-1}(x_v)(x_v^2-1) = (-1)^2 m$ for $v = 1,2,\ldots,m-1$.

(2) $\frac{d}{dx}(x^2-1)U_{m-1}(x)\Big|_{x=\pm 1} = \pm 2U_{m-1}(\pm 1) = \begin{cases} 2m & x = x_0 = -1 \\ (-1)^m 2m & x = x_m = 1 \end{cases}$.

Proof: See Polyá and Szegö (1972).

<u>Lemma 5.3.3</u>: Every polynomial p(x) of degree m can be written as

$$p(x) = \frac{1}{2m} U_{m-1}(x)[p(1)(x+1)+(-1)^m p(-1)(x-1)]$$

$$+ \frac{1}{m} \sum_{v=0}^{m-1} (-1)^v p(x_v) \frac{U_{m-1}(x)(x_v^2-1)}{x-x_v}.$$

<u>Proof</u>: See Polyá and Szegö (1972).

By the definition of Stieltjes transform, we can write

$$\int_{-1}^{1} \frac{d\xi(x)}{z-x} = \frac{1}{\sum_{v=0}^{m} |L_v(z_0)|} \sum_{v=0}^{m} \frac{|L_v(z_0)|}{z-x_v}$$

where $x_v = \cos \frac{v}{m} \pi$ and ξ is the optimal extrapolation design.

It is easily checked that

$$|L_v(z_0)| = \begin{cases} \dfrac{|U_{m-1}(z_0)|(z_0^2-1)}{2m(z_0-1)} & v = 0 \\[2ex] \dfrac{|U_{m-1}(z_0)|(z_0^2-1)}{2m(z_0+1)} & v = 0 \\[2ex] \dfrac{|U_{m-1}(z_0)|(z_0^2-1)}{m(z_0-x_v)} & v = 1,\dots,m-1. \end{cases}$$

For definiteness, we assume $z_0 > 1$. The case $z_0 < -1$ can be treated similarly. By Lemma 5.3.1, Lemma 5.3.2 and Lemma 5.3.3, we can write the Stieltjes transform as

$$\frac{|U_{m-1}(z_0)|(z_0^2-1)}{|T_m(z_0)|}\left[\frac{1}{2m}\frac{1}{z_0-1}\frac{1}{z-1}+\frac{1}{2m}\frac{1}{z_0+1}\frac{1}{z+1}+\frac{1}{m}\sum_{v=1}^{m-1}\frac{1}{(z_0-x_v)(z-x_v)}\right]$$

$$=\frac{|U_{m-1}(z_0)|(z_0^2-1)}{(z-z_0)|T_m(z_0)|}\left[\frac{1}{2m}\left(\frac{1}{z_0-1}+\frac{1}{z_0+1}\right)+\frac{1}{m}\sum_{v=1}^{m-1}\frac{1}{z_0-x_v}-\frac{1}{2m}\left(\frac{1}{z-1}+\frac{1}{z+1}\right)\right.$$

$$\left.-\frac{1}{m}\sum_{v=1}^{m-1}\frac{1}{z-x_v}\right]$$

$$=\frac{|U_{m-1}(z_0)|(z_0^2-1)}{|T_m(z_0)|}\left[\frac{T_m(z_0)}{U_{m-1}(z_0)(z_0^2-1)}-\frac{T_m(z)}{U_{m-1}(z)(z^2-1)}\right]$$

$$=\frac{T_m(z_0)T_m(z)z+[T_{m-1}(z_0)-z_0T_m(z_0)]T_m(z)-T_m(z_0)T_{m-1}(z)}{T_m(z_0)T_m(z)z^2-z_0T_m(z_0)T_m(z)z-T_m(z_0)zT_{m-1}(z)+z_0T_m(z_0)T_{m-1}(z)}$$

Expanding the preceding expression in continued fraction, after a tedious calculation, we obtain (5.3.1).

We can find the canonical moments from the expression (5.3.1). We will give the first three canonical moments in below. By (5.3.1) we see that

$$2\zeta_1-1=\frac{T_{m-1}}{T_m}$$

$$4\zeta_1\zeta_2=\frac{T_{m-1}}{T_m}\left(z_0-\frac{T_{m-1}}{T_m}\right)$$

$$2\zeta_2+2\zeta_3-1=\frac{1}{2}\frac{T_{m-2}}{T_{m-1}}-\frac{T_{m-1}}{T_m}.$$

Hence we obtain

$$p_1=\frac{1}{2}\left(1+\frac{T_{m-1}}{T_m}\right)$$

$$p_2 = \frac{T_{m-1}}{T_m} (z_0 - \frac{T_{m-1}}{T_m}) \frac{1}{1-(\frac{T_{m-1}}{T_m})^2}$$

$$p_3 = \frac{\frac{1}{2} (1 + \frac{1}{2} \frac{T_{m-2}}{T_{m-1}} - \frac{T_{m-1}}{T_m}) - \frac{T_{m-1}}{T_m} (z_0 - \frac{T_{m-1}}{T_m}) \frac{1}{2(1+ \frac{T_{m-1}}{T_m})}}{\frac{1 - z_0 \frac{T_{m-1}}{T_m}}{1 - (\frac{T_{m-1}}{T_m})^2}} .$$

CHAPTER VI

D-OPTIMAL DESIGN FOR MULTIRESPONSE MODEL

6.1 Introduction.

It happens quite often that we have several, say r, regression functions defined on the same interval, say [a,b]. A general discussion of the optimal design in this case can be found in Federov (1972). Läuter (1976) suggested the use of s-optimality criterion in finding optimal designs. In this chapter, we try to find the D-optimal design for the case r = 2 assuming some relationship between the two regression polynomials. In many statistics textbooks on linear models we can find some discussion on the comparison between two regression lines. Given the models

$$y_1 = \beta_{01} + \beta_{11}x + \varepsilon_1$$

$$y_2 = \beta_{02} + \beta_{12}x + \varepsilon_2$$

where ε_1 and ε_2 are uncorrelated normal variables.
Usually there are four models to be considered.

(A) $\beta_{01} \neq \beta_{02}$, $\beta_{11} \neq \beta_{12}$

(B) Parallel regression, $\beta_{11} = \beta_{12}$

(C) Concurrent regression, $\beta_{01} = \beta_{12}$

(D) Coincident regression, $\beta_{01} = \beta_{02}$, $\beta_{11} = \beta_{12}$.

The problem of testing (B) ((C), (D) resp) against (A) is treated under the title comparison of models. The testing of (D) against (B) is usually discussed in the field called analysis of covariance.

In this chapter we will find the D-optimal designs when the regression polynomials are related in different ways. Consider the models

$$\left.\begin{aligned} y_{1i} &= \beta_{01} + \beta_{11}x + \ldots + \beta_{m1}x^m + \varepsilon_{1i} \\ y_{2i} &= \beta_{02} + \beta_{12}x + \ldots + \beta_{n2}x^n + \varepsilon_{2i} \end{aligned}\right\} \qquad (6.1.1)$$

Throughout this chapter we assume all the random variables $\{\varepsilon_{1i}\}$ $\{\varepsilon_{2i}\}$ are independent. Three cases will be considered in the following sections:

(I) The two regression functions have no common parameters. This is related to model (A) above.

(II) Let n = m in 6.1.1. We will consider the case that the highest m-k coefficients of two polynomials are identical. This is a generalization of model (B) above.

(III) Assume n = m in 6.1.1. We will consider the cases:

(i) $\beta_{01} = \beta_{02}$; (ii) $\beta_{01} = \beta_{02}$, $\beta_{11} = \beta_{12}$. This is a direct generalization of model (C) above.

6.2 The D-Optimal Design for Estimating All the Parameters in Two Regression Functions

Consider the model

$$\begin{cases} y_{1i} = \beta_{01} + \beta_{11}x + \ldots + \beta_{m1}x^m + \varepsilon_{1i} \\ y_{2i} = \beta_{02} + \beta_{12}x + \ldots + \beta_{n2}x^n + \varepsilon_{2i}, \quad (n \geq m). \end{cases}$$

We first tackle the problem of finding the D-optimal design for the parameter vector $\beta^T = (\beta_{01},\beta_{11},\ldots,\beta_{m1},\beta_{02},\beta_{12},\ldots,\beta_{n2})$. We have the following theorem.

Theorem 6.2.1: The D-optimal design is given by

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,n$$

$$p_{2i} = \frac{m+n-2i+2}{2m+2n-4i+2} \qquad i = 1,2,\ldots,m$$

$$p_{2i} = \frac{n-i+1}{2n-2i+1} \qquad i = m+1,\ldots,n-1.$$

Proof: The information matrix can be written

$$M(\xi) = \begin{vmatrix} 1 & \mu_1 & \cdots & \mu_m & & & & \\ \mu_1 & \mu_2 & & & & 0 & & \\ \vdots & & & \vdots & & & & \\ \mu_m & \cdots & & \mu_{2m} & & & & \\ & & & & 1 & \mu_1 & \cdots & \mu_n \\ & 0 & & & \mu_1 & \mu_2 & \cdots & \\ & & & & \vdots & & & \vdots \\ & & & & \mu_n & \cdots & & \mu_{2n} \end{vmatrix}$$

The determinant $|M(\xi)|$ can be found to be the product

$$\prod_{i=1}^{m} (\zeta_{2i-1}\zeta_{2i})^{m-i+1} \prod_{i=1}^{n} (\zeta_{2i-1}\zeta_{2i})^{n-i+1}.$$

The result follows immediately.

If we use the following criterion

$$|M_{2m}(\xi)|^{\frac{1}{m+1}} |M_{2n}(\xi)|^{\frac{1}{n+1}} \tag{6.2.1}$$

we will get different results and the resulting design has higher efficiency in estimating the parameters in the lower degree model compared to the design in Theorem 6.2.1.

Theorem 6.2.2: The design that maximizes (6.2.1) has canonical moments

$$p_{2i+1} = \frac{1}{2}, \qquad i = 0,1,\ldots,n-1$$

$$p_{2i} = \frac{2(m+1)(n+1)-i(m+n-1)}{(n+1)(2m+1)+(m+1)(2n+1)-2i(m+n+2)} \quad i = 1,2,\ldots,m$$

$$p_{2i} = \frac{n-i+1}{2n-2i+1} \qquad i = m+1,\ldots,n.$$

Proof: (6.2.1) can be written as

$$[\prod_{i=1}^{m}(\zeta_{2i-1}\zeta_{2i})^{m-i+1}]^{\frac{1}{m+1}}[\prod_{i=1}^{n}(\zeta_{2i-1}\zeta_{2i})^{n-i+1}]^{\frac{1}{n+1}}.$$

By direct calculation, we obtain the desired result.

Example 6.2.1: Let m = 1 and n = 2. The D-optimal design given by Theorem 6.2.1 has canonical moments

$$p_1 = p_3 = \frac{1}{2}$$

$$p_2 = \frac{3}{4}$$

$$p_4 = 1.$$

Following Theorem 6.2.2, we obtain another design with canonical moments:

$$p_1 = p_3 = \frac{1}{2}$$

$$p_2 = \frac{8}{9}$$

$$p_4 = 1 .$$

Suppose m = n and we interest in estimating the vector $(\beta_{01}-\beta_{02}, \beta_{11}-\beta_{12},\ldots,\beta_{m1}-\beta_{m2})^T$. Let

$$A = \begin{pmatrix} 1 & 0 & \ldots & 0 & -1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & -1 & \ldots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & & \vdots \\ 0 & 0 & \ldots & 1 & 0 & & \ldots & 1 \end{pmatrix} = (I_m \vdots -I_m)$$

The covariance matrix of the least squares estimator of $A\beta = (\beta_{01}-\beta_{02},\ldots,\beta_{m1}-\beta_{m2})^T$ is proportional to $AM^{-1}(\xi)A^T$. So it is natural to find a design that will minimize $|AM^{-1}(\xi)A^T|$. In this case, we see

$$|AM^{-1}(\xi)A^T| = \left| [I_m \vdots -I_m] \begin{bmatrix} M_{2m}^{-1}(\xi) & 0 \\ 0 & M_{2m}^{-1}(\xi) \end{bmatrix} \begin{bmatrix} I_m \\ \cdots \\ -I_m \end{bmatrix} \right|$$

$$= |2M_{2m}^{-1}(\xi)|.$$

So the D-optimal design for estimating $A\beta$ is the same as the D-optimal design for $f^T(x) = (1,x,\ldots,x^m)$. We obtain

<u>Theorem 6.2.3</u>: The D-optimal design for estimating $A\beta$ is the same as the D-optimal design for β.

Still assuming m = n, we can find the D-optimal design for estimating $C\beta$, the difference of the s-highest coefficients, where C is given by

$$[0 \vdots I_s \vdots 0 \vdots -I_s].$$

<u>Theorem 6.2.4</u>: The D-optimal design for estimating $C\beta$ is the same as the D_s-optimal design for β.

Proof: It suffices to notice that

$$|CM^{-1}(\xi)C^T| = 2^{m+1}\frac{|M_{2m}(\xi)|}{|M_{11}(\xi)|}$$

where

$$M(\xi) = \begin{bmatrix} M_{2m}(\xi) & \\ & M_{2m}(\xi) \end{bmatrix}$$

and

$$M_{2m}(\xi) = \begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) \end{bmatrix}.$$

6.3 Case II: n = m, the Highest m-k Coefficients of Two Polynomials Are Identical

Consider the model

$$\begin{cases} y_{1i} = \beta_{01} + \ldots + \beta_{k1}x^k + \beta_{k+1}x^{k+1} + \ldots + \beta_m x^m + \varepsilon \\ y_{2i} = \beta_{02} + \ldots + \beta_{k2}x^k + \beta_{k+1}x^{k+1} + \ldots + \beta_m x^m + \varepsilon. \end{cases}$$

The determinant of the information matrix can be written as, assuming X = [0,1],

$$\begin{vmatrix} A & B \\ B^T & C \end{vmatrix}$$

where

$$A = \begin{pmatrix} 1 & 0 & \mu_1 & 0 & \mu_2 & \cdots & \mu_k & 0 \\ 0 & 1 & 0 & \mu_1 & 0 & \cdots & 0 & \mu_k \\ \mu_1 & 0 & \mu_2 & 0 & \mu_3 & \cdots & \mu_{k+1} & 0 \\ \vdots & & & & & & & \vdots \\ 0 & \mu_k & 0 & \mu_{k+1} & 0 & \cdots & 0 & \mu_{2k} \end{pmatrix} \quad B = \begin{pmatrix} \mu_{k+1} & \mu_{k+2} & \cdots & \mu_m \\ \mu_{k+1} & \mu_{k+2} & \cdots & \mu_m \\ \mu_{k+2} & \mu_{k+3} & \cdots & \mu_{m+1} \\ \mu_{k+2} & \mu_{k+3} & \cdots & \mu_{m+1} \\ \vdots & \vdots & & \vdots \\ \mu_{2k+1} & \mu_{2k+2} & \cdots & \mu_{m+k} \\ \mu_{2k+1} & \mu_{2k+2} & \cdots & \mu_{m+k} \end{pmatrix}$$

$$C = \begin{pmatrix} 2\mu_{2k+2} & 2\mu_{2k+3} & \cdots & 2\mu_{m+k+1} \\ 2\mu_{2k+3} & 2\mu_{2k+4} & \cdots & 2\mu_{m+k+2} \\ \vdots & \vdots & & \vdots \\ 2\mu_{m+k+1} & 2\mu_{m+k+2} & \cdots & 2\mu_{2m} \end{pmatrix} .$$

By adding the 2j-1th row (column) to the 2jth row (column), then subtracting $\frac{1}{2}$ times the 2jth row (column) from the 2j-1th row (column), for j = 1,2,...,k+1, we end up with the following determinant

$$\begin{vmatrix} D & E \\ E^T & F \end{vmatrix}$$

where

$$D = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2}\mu_1 & 0 & \frac{1}{2}\mu_2 & \cdots & \frac{1}{2}\mu_k \\ 0 & 2 & 0 & 2\mu_1 & 0 & \cdots & 0 \\ \frac{1}{2}\mu_1 & 0 & \frac{1}{2}\mu_2 & 0 & \frac{1}{2}\mu_3 & \cdots & \frac{1}{2}\mu_{k+1} \\ 0 & 2\mu_1 & 0 & 2\mu_2 & 0 & \cdots & 0 \\ \frac{1}{2}\mu_2 & 0 & \frac{1}{2}\mu_3 & 0 & \frac{1}{2}\mu_4 & \cdots & \frac{1}{2}\mu_{k+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{2}\mu_k & 0 & \frac{1}{2}\mu_{k+1} & 0 & \frac{1}{2}\mu_{k+2} & \cdots & \frac{1}{2}\mu_{2k} \end{pmatrix}$$

$$E = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 2\mu_k & 2\mu_{k+1} & 2\mu_{k+2} & \cdots & 2\mu_m \\ 0 & 0 & 0 & \cdots & 0 \\ 2\mu_{k+1} & 2\mu_{k+2} & 2\mu_{k+3} & \cdots & 2\mu_{m+1} \\ 0 & 0 & 0 & & 0 \\ 2\mu_{k+2} & 2\mu_{k+3} & 2\mu_{k+4} & \cdots & 2\mu_{m+2} \\ \vdots & \vdots & \vdots & & \vdots \\ 2\mu_{2k-1} & 2\mu_{2k} & 2\mu_{2k+1} & \cdots & 2\mu_{m+k-1} \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$F = \begin{pmatrix} 2\mu_{2k} & 2\mu_{2k+1} & \cdots & 2\mu_{m+k} \\ 2\mu_{2k+1} & 2\mu_{2k+2} & \cdots & 2\mu_{m+k+1} \\ \vdots & & & \vdots \\ 2\mu_{m+k} & 2\mu_{m+k+1} & \cdots & 2\mu_{2m} \end{pmatrix} .$$

The above determinant can be further written as

$$(\tfrac{1}{2})^{k+1} \begin{vmatrix} 1 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \vdots & \vdots & & \vdots \\ \mu_k & \mu_{k+1} & \cdots & \mu_{2k} \end{vmatrix} \; 2^{m+1} \begin{vmatrix} 1 & \mu_1 & \cdots & \mu_m \\ \mu_1 & \mu_2 & \cdots & \mu_{m+1} \\ \vdots & \vdots & & \vdots \\ \mu_m & \mu_{m+1} & \cdots & \mu_{2m} \end{vmatrix}$$

$$= 2^{m-k} \prod_{i=1}^{k} (\zeta_{2i-1}\zeta_{2i})^{k-i+1} \prod_{i=1}^{m} (\zeta_{2i-1}\zeta_{2i})^{m-i+1} .$$

To maximize the preceding expression, we have to choose

$$p_{2i} = \frac{m+k-2i+2}{2m+2k-4i+2} \qquad 1 \le i \le k$$

$$p_{2i} = \frac{m-i+1}{2m-2i+1} \qquad k+1 \le i \le m.$$

Thus we have proved the following theorem.

Theorem 6.3.1: The D-optimal design ξ for two general regression functions having the same m-k ($0 \le k \le m$) highest coefficients has canonical moments

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,m-1$$

$$p_{2i} = \frac{m+k-2i+2}{2m+2k-4i+2} \qquad 1 \le i \le k$$

$$p_{2i} = \frac{m-i+1}{2m-2i+1} \qquad k+1 \le i \le m.$$

Remark 6.3.1: If $k = 0$ or $k = m$, we have the D-optimal design.

Example 6.3.1: Let $k = 1$, i.e. the first two coefficients of two polynomials are different while the others are the same. The canonical moments of the D-optimal design are given by

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,m-1$$

$$p_2 = \frac{1+m}{2m}$$

$$p_{2i} = \frac{m-i+1}{2m-2i+1} \qquad i = 2,\ldots,m.$$

Example 6.3.2: For the case $m = k = 2$ i.e. we have the model

$$y_1 = \beta_{01} + \beta_{11}x + \beta_2 x^2 + \varepsilon_1 ,$$

$$y_2 = \beta_{02} + \beta_{12}x + \beta_2 x^2 + \varepsilon_2 .$$

The information matrix can be written

$$\begin{pmatrix} 1 & 0 & \mu_1 & 0 & \mu_2 \\ 0 & 1 & 0 & \mu_1 & \mu_2 \\ \mu_1 & 0 & \mu_2 & 0 & \mu_3 \\ 0 & \mu_1 & 0 & \mu_2 & \mu_3 \\ \mu_2 & \mu_2 & \mu_3 & \mu_3 & 2\mu_4 \end{pmatrix} .$$

The determinant $|M(\xi)| = \text{const } p_1q_1p_2\,(p_1q_1p_2)^2q_2p_3q_3p_4$. It is maximized if

$$p_1 = p_3 = \frac{1}{2}$$

$$p_2 = \frac{3}{4}$$

$$p_4 = 1.$$

6.4 Case III

Consider the model

$$\begin{cases} y_{1i} = \beta_0 + \beta_{11}x + \ldots + \beta_{m1}x^m + \varepsilon_{1i} \\ y_{2i} = \beta_0 + \beta_{12}x + \ldots + \beta_{n2}x^n + \varepsilon_{2i} \end{cases} \qquad n \geq m.$$

The determinant of the information matrix can be written, assuming $X = [0,1]$,

$$\begin{vmatrix} 2 & \mu_1 & \mu_2 & \cdots & \mu_m & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{m+1} & & & \\ \mu_2 & \mu_3 & \mu_4 & \cdots & \mu_{m+2} & & 0 & \\ \vdots & & & & \vdots & & & \\ \mu_m & \mu_{m+1} & \mu_{m+3} & \cdots & \mu_{2m} & & & \\ \mu_1 & & & & & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & & 0 & & & \vdots & & \vdots \\ \mu_n & & & & & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix} \qquad (6.4.1)$$

The above expression is hard to simplify in general. We will find the solutions for several special cases below.

Example 6.4.1: Let $X = [0,1]$ and $m = n = 1$. (6.4.1) becomes

$$\begin{vmatrix} 2 & \mu_1 & \mu_1 \\ \mu_1 & \mu_2 & 0 \\ \mu_1 & 0 & \mu_2 \end{vmatrix} \qquad (6.4.2)$$

which can be further simplified by using the formula

$$\begin{vmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{vmatrix} = |M_{11}|\,|M_{22} - M_{21}M_{11}^{-1}M_{12}| \qquad (6.4.3)$$

(6.4.2) then can be written

$$2\left|\begin{bmatrix}\mu_2 & 0\\ 0 & \mu_2\end{bmatrix} - \frac{1}{2}\begin{bmatrix}\mu_1^2 & \mu_1^2\\ \mu_1^2 & \mu_1^2\end{bmatrix}\right|$$

$$= 2\begin{vmatrix}\mu_2 - \frac{1}{2}\mu_1^2 & -\frac{1}{2}\mu_1^2\\ -\frac{1}{2}\mu_1^2 & \mu_2 - \frac{1}{2}\mu_1^2\end{vmatrix}$$

$$= 2\ \mu_2(\mu_2 - \mu_1^2)$$

$$= 2p_1(p_1+q_1p_2)(p_1q_1p_2).$$

It is easy to see $p_2 = 1$ and $p_1 = \frac{2}{3}$. It should be noted that the D-optimal design in this case is not invariant under a linear transformation on X.

Example 6.4.2: Let $X = [-1,1]$ and $m = n = 1$. The determinant of the information matrix is the same as (6.4.2). If we multiply the first row and first column by -1, the value of the determinant remains unchanged. That means if ξ with support S is D-optimal then so is ξ on -S. Using the inequality

$$|M(\tfrac{1}{2}\,\varepsilon_1 + \tfrac{1}{2}\,\varepsilon_2)| \geq \sqrt{|M(\xi_1)|\,|M(\xi_2)|},$$

we see that ξ is symmetric. Hence $\mu_1 = 0$. (6.4.2) becomes $2\mu_2^2$ which is maximized if $p_2 = 1$. Since ξ is symmetric we have $p_1 = \frac{1}{2}$.

The next three examples also give the D-optimal design on [-1,1].

Example 6.4.3: Let $X = [-1,1]$ and $m = n = 2$. (6.4.1) becomes

$$\begin{vmatrix} 2 & \mu_1 & \mu_1 & \mu_2 & \mu_2 \\ \mu_1 & \mu_2 & 0 & \mu_3 & 0 \\ \mu_1 & 0 & \mu_2 & 0 & \mu_3 \\ \mu_2 & \mu_3 & 0 & \mu_4 & 0 \\ \mu_2 & 0 & \mu_3 & 0 & \mu_4 \end{vmatrix} . \qquad (6.4.3)$$

Using similar method in Example 6.4.2, we see that $\mu_1 = \mu_3 = 0$. (6.4.3) can then be written

$$\mu_2^2 \begin{vmatrix} 2 & \mu_2 & \mu_2 \\ \mu_2 & \mu_4 & 0 \\ \mu_2 & 0 & \mu_4 \end{vmatrix} .$$

The last expression can be further simplified to

$$\mu_2^2 \begin{vmatrix} \mu_4 - \frac{1}{2}\mu_2^2 & -\frac{1}{2}\mu_2^2 \\ -\frac{1}{2}\mu_2^2 & \mu_4 - \frac{1}{2}\mu_2^2 \end{vmatrix}$$

$$= \mu_2^2(\mu_4 - \mu_2^2)\mu_4$$

$$= p_2^2(p_2q_2p_4)p_2(p_2+q_2p_4).$$

The maximization of the last expression gives

$$p_2 = \frac{4}{5}$$

$$p_4 = 1.$$

By symmetry, we have $p_1 = p_3 = \frac{1}{2}$.

Example 6.4.4: Let $X = [-1,1]$ and $m = n = 3$. (6.4.1) becomes

$$\begin{vmatrix} 2 & \mu_1 & \mu_1 & \mu_2 & \mu_2 & \mu_3 & \mu_3 \\ \mu_1 & \mu_2 & 0 & \mu_3 & 0 & \mu_4 & 0 \\ \mu_1 & 0 & \mu_2 & 0 & \mu_3 & 0 & \mu_4 \\ \mu_2 & \mu_3 & 0 & \mu_4 & 0 & \mu_5 & 0 \\ \mu_2 & 0 & \mu_3 & 0 & \mu_4 & 0 & \mu_5 \\ \mu_3 & \mu_4 & 0 & \mu_5 & 0 & \mu_6 & 0 \\ \mu_3 & 0 & \mu_4 & 0 & \mu_5 & 0 & \mu_6 \end{vmatrix} .$$

Using the same technique as in the preceding examples, we can write down the determinant as

$$\begin{vmatrix} \mu_2 & \mu_4 \\ \mu_4 & \mu_6 \end{vmatrix}^2 \begin{vmatrix} 2 & \mu_2 & \mu_2 \\ \mu_2 & \mu_4 & 0 \\ \mu_2 & 0 & \mu_4 \end{vmatrix}$$

$$= p_2^4(q_2p_4q_4p_6)^2\ 2p_2^2q_2p_4(p_2+q_2p_4).$$

By differentiation, we find

$$p_2 = 0.667$$

$$p_4 = 0.616$$

$$p_6 = 1.$$

By symmetry, we know $p_1 = p_3 = p_5 = \frac{1}{2}$.

The last case we consider is to find the D-optimal design if the first two coefficients of the regression functions are equal. Let us consider the following example.

Example 6.4.5: Let $X = [-1,1]$ and $m = n = 2$. The determinant of the information matrix is given by

$$\begin{vmatrix} 2 & 2\mu_1 & \mu_2 & \mu_2 \\ 2\mu_1 & 2\mu_2 & \mu_3 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 & 0 \\ \mu_2 & \mu_3 & 0 & \mu_4 \end{vmatrix} . \qquad (6.4.5)$$

Using the same argument in Example 6.4.2, we see that $\mu_1 = \mu_3 = 0$.

$$2\mu_2 \begin{vmatrix} 2 & \mu_2 & \mu_2 \\ \mu_2 & \mu_4 & 0 \\ \mu_2 & 0 & \mu_4 \end{vmatrix}$$

$$= 2\mu_2 2(\mu_4-\mu_2^2)\mu_4$$

$$= 4p_2^3 q_2 p_4(p_2+q_2p_4)$$

CHAPTER VII

LIMITING DESIGNS, ROBUST-TYPE DESIGNS AND COMPARISON OF EFFICIENCIES

7.1 Some Asymptotic Results

In the last five chapters we have investigated optimal designs for polynomials of degree m under different criteria. In this section we will discuss the limits of these designs when m tends to infinity. A unified approach to finding the limiting design is proposed. Namely, we will identify the limiting designs by their canonical moments. This is valid since the designs we discuss so far are measures on a compact set. We have the following theorem.

Theorem 7.1.1. Let ξ_m denote the weighted D-optimal design for $f^T(x) = \sqrt{w(x)}\,(1,x,\dots,x^m)$, $x \in [0,1]$, where $w(x)$ is one of the following:

(1) $w(x) = 1$

(2) $w(x) = x^{\alpha+1}(1-x)^{\beta+1}$, $\alpha > -1$, $\beta > -1$

(3) $w(x) = x^{\alpha+1}$, $\alpha > -1$

(4) $w(x) = (1-x)^{\beta+1}$, $\beta > -1$.

Then ξ_m converges weakly to ξ_∞, the arc-sine distribution. The density of ξ_∞ is given by $\frac{1}{\pi\sqrt{x(1-x)}}$.

Proof: The case w(x) = 1 was proved by Kiefer and Studden (1976). We will give a similar proof here by using canonical moments. Recall that the canonical moments of the D-optimal design (w(x) = 1) are given by

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,m-1$$

$$p_{2i} = \frac{m-i+1}{2m-2i+1} \qquad i = 1,2,\ldots,m.$$

It is clear that $\lim_{m\to\infty} p_i = \frac{1}{2}$ for all i. Since ξ_∞ is the only distribution with $p_i = \frac{1}{2}$ for all i the result follows.

The proofs of the other cases are similar and so are omitted.

The following six theorems have similar proofs so we will state the theorems only.

Theorem 7.1.2: Let ξ_m denote the weighted D-optimal design for $f^T(x) = \sqrt{w(x)}\,(1,x,\ldots,x^m)$ $x \in [0,1]$, where m is odd and w(x) is one of the following:

(1) $w(x) = [x(1-x)]^\alpha |x - \frac{1}{2}|^\gamma$, α, $\gamma > 0$

(2) $w(x) = |x - \frac{1}{2}|^\gamma$, $\gamma > 0$.

Then ξ_m converges weakly to the arc-sine distribution.

Theorem 7.1.3: (i) The D-optimal design for estimating two regression polynomials converges weakly to the arc-sine distribution if the degree of one of the polynomials tends to infinity.

(ii) The D-optimal design for estimating the difference of the s highest coefficients of two polynomials of the same degree converges weakly to the arc-sine distribution if the degrees of the polynomials tends to infinity.

(iii) The D-optimal design for estimating two polynomials with the same highest k coefficients ($0 \leq k \leq m$) converges weakly to the arc-sine distribution if m tends to infinity.

Theorem 7.1.4: Let ξ_m denote the weighted D_s-optimal design for the s highest coefficients. If the weight function $w(x)$ ($x \in [0,1]$) is one of the following:

(1) $w(x) = 1$

(2) $w(x) = x$

(3) $w(x) = 1-x$

(4) $w(x) = x(1-x)$.

Then ξ_m (s fixed) converges weakly to the arc-sine distribution.

Theorem 7.1.5: Let $X = [-1,1]$ and let ξ_m be the D_s-optimal design ($w(x) = 1$) for the s highest odd (even resp) coefficients. Then ξ_m converges to the arc-sine distribution.

Theorem 7.1.6: Let $X = [-1,1]$ and let ξ_m be the weighted D_s-optimal design for the s highest even coefficients where the weight function is one of the following:

(1) $w(x) = (1-x^2)|x|^2$

(2) $w(x) = |x|^2$.

Then ξ_m converges weakly to the arc-sine distribution.

Theorem 7.1.7: Let ξ_m denote a D-optimal design for $f^T(x) = (1, 2\cos x, 2\sin x, \ldots, 2\cos mx, 2\sin mx)$, $x \in [0, 2\pi]$. Then ξ_m converges weakly to the uniform distribution on the circle.

Theorem 7.1.8: Let the degree of the polynomial m be fixed.

(i) The optimal extrapolation design ($w(x) = 1$) converges weakly to the D_s-optimal design ($s=1$) as $z_0 \to \infty$.

(ii) The optimal extrapolation design ($w(x) = 1$) converges to

$$d\xi(x) = \frac{1}{\pi|z_0 - x|\sqrt{1-x^2}}\,dx \text{ as } m \to \infty.$$

Proof: (i) Using the facts that

$$\lim_{z_0 \to \infty} \frac{T_{n-1}(z_0)}{T_n(z_0)} = 0$$

and

$$\lim_{z_0 \to \infty} z_0 \frac{T_{n-1}(z_0)}{T_n(z_0)} = \frac{1}{2},$$

(5.3.1) becomes

$$\int_{-1}^{1} \frac{d\xi(x)}{z-x} = \frac{1}{z} - \frac{\frac{1}{2}}{z} - \frac{\frac{1}{4}}{z} - \cdots - \frac{\frac{1}{2}}{z}.$$

So (i) is proved.

(ii) Notice that $\lim_{n\to\infty} \frac{T_{n-1}(x_0)}{T_n(x_0)} = \frac{1}{x_0 + \sqrt{x_0^2 - 1}}$ for $x_0 > 1$. So (5.3.1) becomes

$$\int_{-1}^{1} \frac{d\xi(x)}{z-x} = \cfrac{1}{z - \cfrac{1}{x_0+\sqrt{x_0^2-1}}} - \cfrac{\frac{1}{2}\left(1-\left(\frac{1}{x_0+\sqrt{x_0^2-1}}\right)^2\right)}{z + \cfrac{1}{x_0+\sqrt{x_0^2-1}}} - \frac{\frac{1}{4}}{z} - \frac{\frac{1}{4}}{z} - \ldots$$

i.e. $p_1 = \frac{1}{2}(x_0+1 - \sqrt{x_0^2-1})$

$p_2 = p_3 = \ldots = \frac{1}{2}.$

By Example 2.8.3, we know that ξ is absolutely continuous and the density is given by

$$\frac{\sqrt{x_0^2-1}}{\pi(x_0-x)\sqrt{1-x^2}} .$$

If $x_0 < -1$, then $\lim_{n\to\infty} \frac{T_{n-1}(x_0)}{T_n(x_0)} = \frac{1}{x_0-\sqrt{x_0^2-1}}$. So we have

$p_1 = \frac{1}{2}(x_0+1-\sqrt{x_0^2-1})$ and $p_2 = p_3 = \frac{1}{2}$. The density is given by

$$\frac{\sqrt{x_0^2-1}}{\pi(x-x_0)\sqrt{1-x^2}} .$$

There are more limit theorems in the next section. (See Corollary 7.2.1, Corollary 7.2.3).

7.2 Robust-Type Designs

Throughout the last four chapters we have assumed the model is a polynomial of a specified degree m. In many practical situations the exact form of the model is unknown. It happens quite often that the experimenter faces several possible models or in our problem setting, several polynomials of different degrees. We first consider the case that there exists two possible choices: two polynomials of degree r and m respectively with $r < m$. Let us introduce the notations.

Let $X = [-1,1]$ and $f^T(x) = (1,x,\ldots,x^m)$

$$= (f_1(x),f_2(x)), \text{ where } f_1^T(x) = (1,x,\ldots,x^r) \text{ and}$$

$f_2(x) = (x^{r+1},\ldots,x^m)$. We then write $M(\xi)$ as

$$M(\xi) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}.$$

Here M_{11} has size r+1 and M_{22} has size $s = m-r$. By the principle of parismony, we tend to think the model is of degree r but try to guard against the possibility that the highest $s(= m-r)$ coefficients are nonzero. We would like to maximize the determinant $|M_{11}|$ subject to the condition

$$|\Sigma_s(\xi)| = \frac{|M(\xi)|}{|M_{11}(\xi)|} \geq c \qquad (7.2.1)$$

where $\frac{|M(\xi)|}{|M_{11}(\xi)|}$ is proportional to the inverse of the generalized variance of the least squares estimators $\hat{\beta}_{r+1},\ldots,\hat{\beta}_m$. This is the so-called D_{rm} problem (Studden 1981). The solution to this problem is called a D_{rm}-optimal design. A few remarks should be made about the formulation of the D_{rm} problem. Let us give the following definitions first.

Definition 7.2.1: The D-efficiency of a design ξ is given by

$$e_m^D(\xi) = \left(\frac{|M_{2m}(\xi)|}{\max_\eta |M_{2m}(\eta)|}\right)^{\frac{1}{m+1}}$$

where m is the degree of the regression polynomial.

Definition 7.2.2: The D_s-efficiency of a design ξ is given by

$$e_m^{D_s}(\xi) = \left(\frac{|\Sigma_s(\xi)|}{\max_\eta |\Sigma_s(\eta)|}\right)^{\frac{1}{s}}.$$

If we let $c = \rho^s \max_\eta |\Sigma_s(\eta)| (0 \leq \rho \leq 1)$ in (7.2.1), by rearranging the terms, we have

$$\left(\frac{|\Sigma_s(\xi)|}{\max_\eta |\Sigma_s(\xi)|}\right)^{\frac{1}{s}} \geq \rho$$

i.e. $e_m^{D_s}(\xi) \geq \rho$. Since $e_r^D(\xi)$ is proportional to $|M_{11}|$, we therefore have a restatement of the D_{rm} problem: Maximize the D-efficiency of degree r subject to the condition that the D_s-efficiency is larger than or equal to ρ. The problem is hard to solve in Stigler's original approach and he gave the solution for the D_{12} problem only. By using canonical moments, Studden (1981) can give solutions for D_{1m} and D_{2m} problems. (See Stigler (1971).)

Theorem 7.2.1: The D_{1m} optimal design has canonical moments

$$p_{2i-1} = \frac{1}{2} \qquad i = 1,2,\ldots,m$$

$$p_2 = \frac{1+\sqrt{1-\rho}}{2}$$

$$p_{2i} = \frac{m-i+1}{2m-2i+1} \quad i = 2,3,\ldots,m-1$$

$$p_{2m} = 1.$$

Proof: See Studden (1982).

Corollary 7.2.1: D_{1m} optimal design converges weakly to $\xi_{1\infty}$ having canonical moments $p_2 = \frac{1+\sqrt{1-\rho}}{2}$ and $p_i = \frac{1}{2}$ for $i \neq 2$. $\xi_{1\infty}$ is absolutely continuous with density

$$\frac{1}{\pi\sqrt{1-x^2}} \frac{\rho}{(1+\sqrt{1-\rho})^2 - 4\sqrt{1-\rho}\,x^2} .$$

Proof: The limits of the canonical moments are easy to see. The density was calculated in Example 2.8.4.

Theorem 7.2.2: The D_{2m} optimal design has canonical moments

$$p_{2i-1} = \frac{1}{2} \qquad i = 1,2,\ldots,m$$

$$p_{2i} = \frac{m-i+1}{2m-2i+1} \quad i = 3,4,\ldots,m-1$$

$$p_{2m} = 1.$$

p_2 is the root of the equation

$$\rho(1-2p_2)^2 + 16(2-3p_2)(p_2q_2 - \tfrac{\rho}{4})(p_2-1) = 0.$$

p_4 is given by

$$p_4 = \frac{1}{2}\,[1 + (1 - \frac{\rho}{4p_2q_2})^{\frac{1}{2}}].$$

Proof: See Studden (1982).

Notice that the D_{1m} optimal design has the same canonical moments except for p_2 as the D-optimal design. The D_{2m} optimal design has the same canonical moments as the D-optimal design except for p_2 and p_4. The same pattern holds for $r \geq 3$. It is also clear from these examples that the D_{rm} optimal designs for $m \geq r$ share the same first r even canonical moments. We see that the explicit solution of p_2 and p_4 cannot be found in the D_{2m} problem. We can expect some complexity in the D_{rm}-optimal design as r grows larger.

In using the D_{rm} design, we will first test the hypothesis $\beta_{r+1} = \beta_{r+2} = \ldots = 0$ versus the alternative H_1: not all $\beta_i = 0$, $i = r+1,\ldots,m$; if we accept the hypothesis, fit the simpler model, if we reject the hypothesis, fit the complete model. It seems interesting to investigate the relation between design and testing, or the power of the test.

In testing the hypothesis $H\beta = 0$, we know the power of the test is an increasing function of the non-centrality parameter λ which is given by

$$\lambda = \frac{1}{\sigma^2} \beta^T H^T [H^T (x^T x)^{-1} H]^{-1} H\beta .$$

In our case, we have $H = [0 \vdots I_s]$ where I_s is the identity matrix of size s and $[H^T(x^Tx)^{-1}H]^{-1} = M_{22} - M_{21}M_{11}^{-1} M_{12}$. Let $\beta^T = (\underset{\sim}{\beta}_1^T, \underset{\sim}{\beta}_2^T)$, where $\underset{\sim}{\beta}_1$ is the (r+1)-vector and $\underset{\sim}{\beta}_2$ is the s-vector. Then we have

$$\lambda = \frac{n}{\sigma^2} \underset{\sim}{\beta}_2^T (M_{22} - M_{21} M_{11}^{-1} M_{12}) \underset{\sim}{\beta}_2 .$$

It should be noted that λ is dependent on the regression interval. In the case s = 1 and X = [-1,1], we see that

$$\lambda = \frac{n}{\sigma^2}\, 2^{2(r+1)} \prod_{i=1}^{2r+1} (p_i q_i) p_{2r+2} \beta_{r+1}^2$$

λ is maximized iff $p_1 = p_2 = \ldots = p_{2r+1} = \frac{1}{2}$ and $p_{2r+2} = 1$. Note that in this case the condition $|e_m^{D_s}(\xi)| \geq \rho$ is equivalent to $\lambda \geq$ some constant. In the case s = 2, X = [-1,1] and assuming symmetry of the design, we see that

$$\lambda = \frac{n}{\sigma^2} [\beta_{r+1}, \beta_{r+2}] \begin{bmatrix} p_2q_2p_4\cdots p_{2r+2} & 0 \\ 0 & p_2q_2p_4\cdots p_{2r+4} \end{bmatrix} \begin{bmatrix} \beta_{r+1} \\ \beta_{r+2} \end{bmatrix} .$$

For each λ, the above expression describes an ellipse in the $(\beta_{r+1}, \beta_{r+2})$ plane. To ensure the test has the desired power, we have to choose the sample size n and the design so that the acceptance region (sphere) contains the above ellipse. Notice that the longer axis will correspond to the minimum eigenvalue of the matrix $M_{22}-M_{21}M_{11}^{-1}M_{12}$. It is clear that the design that maximizes the minimum eigenvalue will give the higher power. In this case the minimum eigenvalue is $p_1q_1p_2q_2\cdots p_{2r+4}$ and it is maximized iff $p_2 = p_4 = \ldots = p_{2r+2} = \frac{1}{2}$ and $p_{2r+4} = 1$. All odd canonical moments are $\frac{1}{2}$ by symmetry. We will compare efficiencies of different designs in the next section. In the following, we will use the same example suggested by Stigler (1971) to see how the D_{13}-optimal design is used in practice.

Suppose we wish to design an experiment to estimate the interference effect (measured by induced voltage) of a certain power line on phone calls as a function of distance. Based on past experience with similar situations, we know that the log of the induced voltage behaves approximately as a linear function of the distance from the power line, and that the errors of measurement are independently normally distributed random variables with a common variance. Let us further simplify the problem by transforming (linearly) the scale of the distance measurements (x) so that our range of interest is from x = -1 to x = 1, and by assuming that the common variance of the errors is known.

Our aim here is to design the experiment in such a way that if the true relationship differs significantly from a linear one we will detect the difference with high probability. Also we want to have high D-efficiency in estimating the model we choose after the testing.

As a measure of departure from linearity, we could consider the non-centrality parameter, i.e.

$$\lambda = \frac{n}{\sigma^2} [\beta_2, \beta_3] \begin{bmatrix} p_2 q_2 p_4 & \\ & p_2 q_2 p_4 q_4 p_6 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} .$$

For the case under consideration we might define "does not differ significantly from a linear relationship" to mean that

$$\frac{(\beta_2)^2}{\sigma^2} + \frac{(\beta_3)^2}{\sigma^2} \leq (0.25)^2.$$

The number 0.25 was chosen since, for the purposes of prediction to which the model will be put, the presence of $\beta_2 x^2 + \beta_3 x^3$ with $|\beta_2| \leq 0.25\sigma$ and $|\beta_3| \leq 0.25\sigma$ would be largely "washed out" by random error.

Let us decide to follow the procedure "test the hypothesis $H_0 = \beta_2 = \beta_3 = 0$ versus the alternative $H_1 = \beta_2 \neq 0$ or $\beta_3 \neq 0$ at the significance level $\alpha = 0.10$ with power of 0.70 at the alternative $|\beta_2| = 0.25\sigma$ (β_3=0) or $|\beta_3| = 0.25\sigma$ (β_2=0). For the above significance level and power, the non-centrality parameter is 5.974 (see Selected Tables in Math. Stat. Vol. 1, IMS). For a given value of the non-centrality parameter and design, we can find the sample size that will achieve the required power and significance level. In this case, we have

$$\lambda = \frac{n}{\sigma^2} (p_2 q_2 p_4 \beta_2^2 + p_2 q_2 p_4 q_4 p_6 \beta_3^2).$$

By Theorem 7.2.1, we see that $p_2 = \frac{1+\sqrt{1-\rho}}{2}$, $p_4 = \frac{2}{3}$ and $p_6 = 1$. Given that $|\frac{\beta_3}{\sigma}| \leq \frac{1}{4}$, we find

$$n\rho \geq 1720.51.$$

That means the sample size should be at least 1121. If we take $\rho = 0.9$, the sample size should be at least 1912. If we choose $\rho = 0.72$, we find n should be at least 2390.

7.3 The Comparison of Designs

We will compare different designs under different criteria. Let us introduce the following designs first.

(1) The minimum bias design ξ_{BD}. Box and Draper (1959) suggested the design which has its first 2r+s moments agree with the first 2r+s moments of the uniform distribution on [-1,1]. This design is supposed to have the bias (that caused by the s extra terms) minimized. If s = 1, the minimum bias design is given by

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,r$$

$$p_{2i} = \frac{i}{2i+1} \qquad i = 1,2,\ldots,r$$

$$p_{2r+2} = 1.$$

If s = 2, the minimum bias design is given by

$$p_{2i+1} = \frac{1}{2} \qquad i = 0,1,\ldots,r+1$$

$$p_{2i} = \frac{i}{2i+1} \qquad i = 1,2,\ldots,r+1$$

$$p_{2r+4} = 1.$$

From the theory of moments, the design ξ_{BD} has its support rested on the zeros of $(x^2-1)P_m'(x) = 0$, where $P_m'(x)$ is the derivative of the m-th Legendre polynomial and the weight attached to x_i is given by

$$\frac{1}{m(m+1)P_m^2(x_i)} .$$

(2) The discrete uniform design ξ_{un} on n equally spaced points on $[-1,1]$, i.e. $-1 = x_1 < x_2 <\ldots< x_n = 1$. The canonical moments are given by (See Example 2.3.1).

$$p_{2i+1} = \frac{1}{2} \qquad i = 1,2,\ldots,n-2$$

$$p_{2i} = \frac{i}{2i+1}\,\frac{n+i}{n-1} \qquad i = 1,2,\ldots,n-2$$

$$p_{2n-2} = 1.$$

(3) B-optimal design

Läuter (1974) suggested using the criterion function

$$|M(\xi)|^{w_1}|M_{11}(\xi)|^{w_2}, \text{ or}$$

$$w_1 \ln|M(\xi)| + w_2 \ln|M_{11}(\xi)|$$

where w_1, w_2 are nonnegative constants which may be chosen to get a desired relation of orders between the different factors. If we take $w_1 = \frac{1}{m+1}$ and $w_2 = \frac{1}{r+1}$. We call the resulting design the B_{rm}-optimal design. So B_{rm}-optimal design can be interpreted as the design that maximizes the product of efficiencies of the models. It is easy to prove the following theorem.

<u>Theorem 7.3.1</u>: The B_{rm}-optimal design has canonical moments

$$p_{2i-1} = \frac{1}{2}, \qquad i = 1,2,\ldots,m$$

$$p_{2i} = \frac{2mr+2(m+r-i)-(m+r)i}{4mr+3m+3r-4i-2i(m+r)+2}, \quad i = 1,2,\ldots,r$$

$$p_{2i} = \frac{m-i+1}{2m-2i+1}, \qquad i = r+1,\ldots,m.$$

Corollary 7.3.1: Let ξ be the limiting design of B_{rm}-optimal design when (r fixed) m tends to infinity. Then ξ has canonical moments

$$p_{2i-1} = \frac{1}{2}, \qquad i = 1,2,\ldots,m$$

$$p_{2i} = \frac{2r+2-i}{4r+3-2i}, \qquad i = 1,2,\ldots,r$$

$$p_{2i} = \frac{1}{2}, \qquad i = r+1,\ldots,m.$$

Next we will introduce other ways to compare the performances of different designs.

Definition 7.3.1: The I_σ-efficiency of a design ξ is given by

$$e_m^{I_\sigma}(\xi) = \frac{\min_\mu \operatorname{tr} M^{-1}(\mu)M(\sigma)}{\operatorname{tr} M^{-1}(\xi)M(\sigma)} .$$

It is difficult in general to find $\min_\mu \operatorname{tr} M^{-1}(\mu)M(\sigma)$. Some numerical work has been done in Studden (1977) for the case $d\sigma(x) = \frac{1}{2}\,dx$. The denominator can be expressed in terms of canonical moments as we have shown in Chapter V. For the first four integrated variances $\operatorname{tr} M^{-1}(\xi)M(u)$ where u denotes the uniform measure

$$v_1(\xi,u) = 1 + \frac{1}{3p_2}$$

$$v_2(\xi,u) = v_1(\xi,u) + \frac{1}{p_2q_2p_4}\left[\frac{4}{45} + \left(\frac{1}{3} - p_2\right)^2\right]$$

$$v_3(\xi,u) = v_2(\xi,u) + \frac{1}{p_2q_2p_4q_4p_6}\left[\frac{4}{175} + \frac{1}{3}\left(\frac{2}{5} - q_2q_4\right)^2\right]$$

$$v_4(\xi,u) = v_3(\xi,u) + \frac{1}{p_2q_2p_4q_4p_6q_6p_8}\,[(\frac{8}{105})^2 + \frac{4}{135}$$

$$(-\frac{1}{7} - q_4p_6 + q_2q_4)^2 + [\frac{1}{3}(q_2q_4 - \frac{2}{5}) + q_4p_6(q_2 - \frac{1}{3})]^2.$$

We also write the first four D-efficiencies down because we will use them for comparison.

$$e_1^D(\xi) = \sqrt{p_2}$$

$$e_2^D(\xi) = 3(\frac{p_2^2q_2p_4}{4})^{\frac{1}{3}}$$

$$e_3^D(\xi) = \frac{1}{2}[5^5p_2^3(q_2p_4)^2q_4p_6]^{\frac{1}{4}}$$

$$e_4^D(\xi) = \frac{35}{4}[\frac{49}{108}\,p_2^4(q_2p_4)^3(q_4+p_6)^2q_6p_8]^{\frac{1}{5}}.$$

Although it is not our main concern here to discuss the testing power, we will include a comparison of powers of different designs in terms of their sample sizes. For example, in case $r = 1$ and $s = 2$, we equate two noncentrality parameters, i.e.

$$np_2q_2p_4q_4p_6 = n'p_2'q_2'p_4'q_4'p_6'$$

$$\frac{n}{n'} = \frac{p_2'q_2'p_4'q_4'p_6'}{p_2q_2p_4q_4p_6}.$$

Here n and n' denote the sample size. If we choose $p_2 = p_4 = \frac{1}{2}$ and $p_6 = 1$, i.e. the design with maximum power, then n will be the smallest possible sample size for the given power and significance level. Thus the ratio $\frac{n}{n'}$ will measure the efficiency of the design in testing. In this case, the ratio $\frac{n}{n'}$ can be written as

$$\frac{n}{n'} = 16p_2'q_2'p_4'q_4'p_6'\;.$$

If r = 2 and s = 2, we have, by equating their minimum eigenvalues of the matrix Σ_s,

$$\frac{n}{n'} = \frac{p_2'q_2'p_4'q_4'p_6'q_6'p_8'}{p_2q_2p_4q_4p_6q_6p_8} .$$

To achieve the highest power we just need to set $p_2 = p_4 = p_6 = \frac{1}{2}$ and $p_8 = 1$. Thus we have

$$\frac{n}{n'} = 64p_2'q_2'p_4'q_4'p_6'q_6'p_8.$$

The ratio $\frac{n}{n'}$ again indicates the efficiency of the design in testing.

Before we give the canonical moments of the designs that we are going to compare we first introduce the following notations:

1. ξ_{BD} denotes the minimum bias design.
2. ξ_{un} denotes the discrete uniform design on n equally spaced points.
3. ξ_{rm} denotes the D_{rm}-optimal design.
4. ξ_B denotes the B-optimal design.
5. ξ_m denotes the D-optimal design.

In case r = 1 and s = 2 the canonical moments of above designs are given in the following table. Notice that all odd canonical moments are equal to $\frac{1}{2}$.

	p_2	p_4	p_6
ξ_{BD}	0.333	0.400	1
ξ_{u4}	0.556	0.800	1
ξ_{u10}	0.407	0.533	0.619
ξ_{u100}	0.340	0.412	0.446
ξ_{u300}	0.336	0.404	0.434
$\xi_{13}(=0.9)$	0.658	0.667	1
$\xi_{13}(=0.72)$	0.765	0.667	1
ξ_B	0.714	0.667	1
ξ_3	0.600	0.667	1

For the case r = 2, s = 2, the canonical moments are given by:

	p_2	p_4	p_6	p_8
ξ_{BD}	0.333	0.400	0.428	1
ξ_{u5}	0.500	0.700	0.857	1
ξ_{u10}	0.407	0.533	0.619	0.691
ξ_{u100}	0.34	0.412	0.446	0.467
$\xi_{24}(\rho=0.7)$	0.625	0.752	0.667	1
ξ_B	0.611	0.700	0.667	1
ξ_4	0.571	0.600	0.667	1

Table 7.3.1. The efficiencies of various designs for the case r = 1 and s = 2.

ξ	$e_1^D(\xi)$	$e_2^D(\xi)$	$e_3^D(\xi)$	$e_1^{I_\sigma}(\xi)$	$e_2^{I_\sigma}(\xi)$	$e_3^{I_\sigma}(\xi)$	$\frac{n}{n'}$
ξ_{BD}	0.577	0.585	0.745	0.667	0.710	0.872	0.853
ξ_{u5}	0.746	0.905	0.959	0.833	0.926	0.809	0.632
ξ_{u10}	0.638	0.707	0.785	0.733	0.835	0.905	0.595
ξ_{u100}	0.583	0.597	0.621	0.673	0.724	0.769	0.388
ξ_{u300}	0.580	0.589	0.609	0.669	0.715	0.755	0.373
$\xi_{13}(\rho=0.9)$	0.811	0.874	0.991	0.885	0.760	0.786	0.800
$\xi_{13}(\rho=0.72)$	0.875	0.852	0.920	0.929	0.571	0.578	0.639
ξ_B	0.845	0.869	0.963	0.909	0.669	0.684	0.726
ξ_3	0.775	0.866	1	0.857	0.833	0.872	0.853

Table 7.3.2. The efficiencies of various designs for the case r = 2 and s = 2.

ξ	$e_2^D(\xi)$	$e_3^D(\xi)$	$e_4^D(\xi)$	$e_2^{I_\sigma}(\xi)$	$e_3^{I_\sigma}(\xi)$	$e_4^{I_\sigma}(\xi)$	$\frac{n}{n'}$
ε_{BD}	0.585	0.603	0.728	0.710	0.748	0.870	0.835
ε_{u5}	0.839	0.936	0.900	0.913	0.905	0.858	0.412
ε_{u10}	0.707	0.785	0.850	0.835	0.905	0.974	0.627
ε_{u100}	0.597	0.621	0.645	0.724	0.769	0.782	0.401
$\varepsilon_{24}(\rho=0.7)$	0.906	0.890	0.938	0.845	0.682	0.747	0.621
ε_B	0.882	0.902	0.973	0.837	0.741	0.817	0.710
ε_4	0.828	0.895	1	0.828	0.831	0.925	0.836

From Table 7.2.1 and Table 7.2.2, it can be seen that ξ_{BD} does very poorly in D-efficiencies and it does a little better in I_σ-efficiencies. The discrete uniform designs perform quite well when the number of design points, say n, is closed to m+1. When n gets larger, both the efficiencies and testing power are decreasing. Both ξ_{13} and ξ_{24} have high efficiencies in D-optimality. The D-optimal designs ξ_m (m = 3,4) performs very well in different criteria except in the lower degree model. On the whole, it seems that ξ_{13} and ξ_{24} perform quite well no matter which model is used. The ξ_B design behaves similar to ξ_{13} and ξ_{24}.

Below we will discuss the asymptotic efficiency of the various designs we mentioned above. Let us investigate the limit of D-efficiency first.

Theorem 7.3.2: Let ξ_∞ denote the arc-sine distribution. Then

$$\lim_{m\to\infty} e_m^D(\xi_\infty) = 1.$$

Proof: See Kiefer and Studden (1976).

The following theorem gives the limit of the D-efficiency of a design ξ satisfying certain properties.

Theorem 7.3.3: Let ξ be a probability measure with canonical moments p_i, i = 1,2,..., such that $\sum_{i=1}^{\infty}(p_i - \frac{1}{2})^2 < \infty$. Suppose further that $\xi'(x) = w(x)$. Then we have

$$\lim_{m\to\infty} e_m^D(\xi) = \frac{\pi}{2}\exp\left(\frac{1}{\pi}\int_{-1}^{1}\frac{\log w(x)}{\sqrt{1-x^2}}\,dx\right).$$

Proof: By Corollary 2.8.6, we see that

$$\lim_{m\to\infty} \rho_m^D(\xi) = \frac{\lim\limits_{m\to\infty} 2^{m+1}|M_{2m}(\xi)|^{\frac{1}{m+1}}}{\lim\limits_{m\to\infty} 2^{m+1}|M_{2m}(\xi_m)|^{\frac{1}{m+1}}} = \frac{2\pi\exp\{\frac{1}{\pi}\int_{-1}^{1} \frac{\log w(x)}{\sqrt{1-x^2}}\, dx\}}{\lim\limits_{m\to\infty} 2^{m+1}|M_{2m}(\xi)|^{\frac{1}{m+1}}}.$$

By Theorem 7.2.5, the denominator can be replaced by

$$\lim_{m\to\infty} 2^{m+1}|M_{2m}(\xi_m)|^{\frac{1}{m+1}},$$

which is, by Corollary 2.8.6,

$$2\pi\exp\{\frac{1}{\pi}\int_{-1}^{1} \frac{\log \frac{1}{\pi\sqrt{1-x^2}}}{\sqrt{1-x^2}}\, dx\}$$

$$= 2\pi\exp\{-\log \pi - \frac{1}{\pi}\int_0^{\pi} \log \sin\theta d\theta\}$$

$$= 4.$$

So we have

$$\lim_{m\to\infty} e_m^D(\xi) = \frac{\pi}{2} \exp\{\frac{1}{\pi}\int_{-1}^{1} \frac{\log w(x)}{\sqrt{1-x^2}}\, dx\}.$$

The next theorem was proved by Kiefer and Studden (1976). We first have a lemma.

Lemma 7.3.4: Let f and g be non-negative and integrable functions with respect to a measure μ and s be the region in which $f(x) > 0$. If $\int_s (f-g)d\mu \geq 0$, then

$$\int_S f(x) \log g(x) d\mu$$

is maximized uniquely by taking $g(x) = f(x)$ (μ a.e.).

Proof: See Rao (1973).

Theorem 7.3.4: $\lim_{m\to\infty} e_m^D(\xi)$ is maximized by taking $w(x) = \frac{1}{\pi\sqrt{1-x^2}}$ where ξ satisfies the conditions mentioned in Theorem 7.2.6.

Proof: Directly follows from Lemma 7.2.7.

In case $\xi = \xi_u$, the uniform measure on the interval [-1,1], we have the following theorem.

Theorem 7.3.5: $\lim_{m\to\infty} e_m^D(\xi_u) = \frac{\pi}{4}$.

Proof: By Theorem 7.2.6, we have

$$\lim_{m\to\infty} e_m^D(\xi_u) = \frac{\pi}{2} \exp\{\frac{1}{\pi} \int_{-1}^{1} \frac{\log \frac{1}{2}}{\sqrt{1-x^2}} dx\}$$

$$= \frac{\pi}{4}.$$

Theorem 7.3.6: Let ξ_{BD} denote the minimum bias design with all its first 2m-1 canonical moments identical with those of the uniform measure and $p_{2m} = 1$. Then

$$\lim_{m\to\infty} e_m^D(\xi_{BD}) = \frac{\pi}{4}.$$

Proof: Notice that

$$\left(\frac{|M_{2m}(\xi_{BD})|}{|M_{2m}(\xi_u)|}\right)^{\frac{1}{m+1}} = \left(\frac{1}{p_{2m}}\right)^{\frac{1}{m+1}} = \left(\frac{2m+1}{m}\right)^{\frac{1}{m+1}}.$$

Hence we have

$$\lim_{m\to\infty} \left(\frac{|M_{2m}(\xi_{BD})|}{|M_{2m}(\xi_u)|}\right)^{\frac{1}{m+1}} = 1$$

and the result follows.

Theorem 7.3.7: Let ξ_{un} denote the design with equal mass on n points. Then

$$\lim_{m\to\infty} \lim_{n\to\infty} e_m^D(\xi_{un}) = \frac{\pi}{4}.$$

Proof: Observe that ξ_{un} converges weakly to the uniform measure as $n \to \infty$. For fixed m, we have

$$\lim_{n\to\infty} e_m^D(\xi_{un}) = e_m^D(\xi_u).$$

Hence the result follows.

Theorem 7.3.8: Let ξ_{1m} denote the optimal design for D_{1m}-problem. Then, we have

$$\lim_{m\to\infty} e_m^D(\xi_{1m}) = \lim_{m\to\infty} e_m^D(\xi_{1\infty}) = \rho.$$

Proof: Notice that ξ_{1m} has the same canonical moments as the D-optimal design except p_2. $\xi_{1\infty}$ has the same canonical moments as ξ_∞, the arc-sine law, except p_2. But p_2 for ξ_{1m} is the same p_2 for $\xi_{1\infty}$. By comparing with Theorem 7.2.5, we have

$$\lim_{m\to\infty}\left(\frac{|M_{2m}(\xi_{1m})|}{|M_{2m}(\xi_{1\infty})|}\right)^{\frac{1}{m+1}} = 1.$$

Hence the first part of the theorem is proved.

By Theorem 7.2.6, we have

$$\lim_{m\to\infty} e_m^D(\xi_{1\infty}) = \frac{\pi}{2}\exp\{\frac{1}{\pi}\int_{-1}^{1}\frac{\log(\frac{\rho}{\pi\sqrt{1-x^2}[(1+\sqrt{1-\rho})^2-4\sqrt{1-\rho}x^2]})}{\sqrt{1-x^2}}dx\}$$

$$= \exp\{\frac{1}{\pi}\int_0^{\pi}(\log\rho-\log[(1+\sqrt{1-\rho})^2-4\sqrt{1-\rho}\cos^2\theta])d\theta.$$

After some calculations we find

$$\lim_{m\to\infty} e_m^D(\xi_{1\infty}) = \rho.$$

The following theorem gives the form of the asymptotic D_s-efficiency.

Theorem 7.3.9: Let ξ be the measure with canonical moments p_i, $i = 1,2,\ldots$, such that $\sum_{i=1}^{\infty}(p_i - \frac{1}{2})^2 < \infty$. Then

$$\lim_{m\to\infty} 2^{2m-s+2} e_m^{D_s}(\xi) = 2\pi\exp\{\frac{1}{\pi}\int_{-1}^{1}\frac{\log w(x)}{\sqrt{1-x^2}}dx\}$$

where $w(x) = \xi'(x)$, the derivative of ξ.

Proof: By Theorem 2.8.7a, we see that

$$\lim_{m\to\infty}\left[2^{2m+1}\frac{|M_{2m}(\xi)|}{|M_{2m-2}(\xi)|}\,2^{2m-1}\frac{|M_{2m-2}(\xi)|}{|M_{2m-4}(\xi)|}\cdots 2^{2(m-s+1)+1}\frac{|M_{2(m-s+1)}(\xi)|}{|M_{2(m-s+1)}(\xi)|}\right]^{\frac{1}{2}}$$

$$= 2\pi \exp\{\frac{1}{\pi} \int_{-1}^{1} \frac{\log w(x)}{\sqrt{1-x^2}} dx\}.$$

By simplifying the left-hand side, we obtain the desired result.

Use the same reasoning in Theorem 7.2.7, we have the following theorem.

Theorem 7.3.10: $\lim_{m\to\infty} 2^{2m-s+2} e_m^{D_2^2}(\xi)$ is uniquely maximized by taking ξ to be arc-sine.

Proof: Use Lemma 7.2.7.

Definition 7.3 .2: The G-efficiency of a design ξ is given by

$$e_m^G(\xi) = \frac{m+1}{\max_x d(x,\xi)}.$$

Theorem 7.3.11: $\lim e_m^G(\xi_\infty) = \frac{1}{2}.$

Proof: See Kiefer and Studden (1976).

The following theorem shows that the uniform design has G-efficiency equal to 0.

Theorem 7.3.12: $\lim_{m\to\infty} e_m^G(\xi_u) = 0.$

Proof: From Guest (1958) it is known that

$$d_m(x,\xi_u) \approx (m+1)^2 p_k^2(x) - (x^2-1)[p_k'(x)]^2$$

when m is large. Here $p_k(x)$ denotes the k-th Legendre polynomial. Guest (1958) has also shown that

$$\max_x d_m(x,\xi_u) = d_m(\pm 1, \xi_u) = (m+1)^2.$$

Hence we have

$$\lim_{m\to\infty} e_m^G(\xi_u) = \lim_{m\to\infty} \frac{1}{m+1} = 0.$$

For $x \in (-1,1)$, we have the following comparison between ξ_∞ and ξ_u.

Theorem 7.3.13: For $x \in (-1,1)$, we have

$$\lim_{m\to\infty} \frac{d_m(x,\xi_\infty)}{d_m(x,\xi_u)} = \frac{\pi}{2}\sqrt{1-x^2}.$$

Proof: Use Theorem 2.8.8.

The result shows that ξ_∞ performs better than ξ_u in the region $|x| \geq 0.7718$.

BIBLIOGRAPHY

Atwood, C. L. (1969). Optimal and efficient designs of experiments. Ann. Math. Statist., 40, 1570-1602.

Box, G.E.P. and Draper, N. R. (1959). A basis for the selection of a response surface design, JASA, 54, 622-654.

Brezinski, C. (1980). Padé-type approximation and general orthogonal polynomials. ISNM 50 Birkhäuser Verlag.

Broth, D. M. (1975). A total entropy criterion for the dual problem of model discrimination and parameter estimation. J. Roy. Statist. Soc. Ser. B, 37, 77-87.

Chihara, T. S. (1978). An introduction to orthogonal polynomials. Gordon and Breach.

Cook, R. D. and Nachtsheim, C. J. (1982). Model robust, linear-optimal designs. Technometrics, 24, 49-54.

Farrell, R. M., Kiefer, J. and Walbran, A. (1967). Optimum Multivariate Designs. Proc. 5th Berkeley Symp. Math. Statist. Prob. 1, 113-138.

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

Freud, G. (1971). Orthogonal Polynomials. Pergamon Press, New York.

Geronimus, Ya L. (1948). Polynomial orthogonal on a circle and their applications. Translated by R. P. Boas. Translation Series 1, Vol. 3, Am. Math. Soc., Providence, Rhode Island.

Geronimus, Ya L. (1950). On asymptotic properties of polynomials which are orthogonal on the circle, and on certain properties of positive harmonic functions. Translated by F. V. Atkinson. Translation Series 1, Vol. 3, Am. Math. Soc., Providence, Rhode Island.

Geronimus, Ya L. (1961a). Orthogonal Polynomials. Consultants Bureau, New York.

Geronimus, Ya L. (1961b). Orthogonal Polynomials. Am. Math. Soc. Translation Ser. 2, Vol. 108.

Gantmacher, F. R. (1959). The Theory of Matrices. Chelsea Publishing Company.

Grenander, U. and Szegö, G. (1958). Toeplitz Forms and Their Applications. University of California Press.

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

Henrici, P. (1973). Applied and Computational Complex Analysis. John Wiley and Sons.

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

Jones, W. B. and Thron, W. J. (1980). Continued Fractions Analytic Theory and Applications. Encyclopedia of Mathematics and Its Applications, Vol. 11.

Karlin, S. and McGregor, J. (1961). The Hahn polynomials, formulas and an application. Scripta Math., 26, 33-46.

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

Karlin, S. and Studden, W. J. (1966b). Optimal experimental designs. Ann. Math. Statist., 37, 783-815.

Kiefer, J. (1959). Optimum experimental designs. J. Roy. Statist. Soc. Ser. B, 21, 272-319.

Kiefer, J. (1961a). Optimum designs in regression problems II. Ann. Math. Statist., 32, 298-325.

Kiefer, J. (1961b). Optimum experimental designs V with applications to systematic and rotatable designs. Proc. 4th Berkeley Symp. Math. Statist. Prob. 1, 381-405.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximation theory). Ann. Statist., 2, 849-879.

Kiefer, J. (1975). Optimal design: Variation in structure and performance under change of criterion. Biometrika 62, 277-288.

Kiefer, J. and Studden, W. J. (1976). Optimal designs for large degree polynomial regression. Ann. Statist. 4, 1113-1123.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. Ann. Math. Statist. 30, 271-294.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canad. J. Math., 12, 363-366.

Krein, M. G. and Nudelman, A. A. (1977). The Markov Moment Problem and Extremal Problems. Translated by Louvish, D., Am. Math. Soc., Providence, Rhode Island.

Läuter, E. (1974). Experimental planning in a class of models. Math. Oper. Statist., 5, 673-708.

Läuter, E. (1976). Optimal multipurpose designs for regression models. Math. Oper. Statist. 7, 51-67.

Lesky, P. (1962). Erganzung zur Übersetzung der Klessischen orthogonalen Polynome in die Differenzenrechnung. Monat. Math., 66, 431-434.

Marcus, M. B. and Sacks, J. (1976). Robust designs for regression problems. Statistical Decision Theory and Related Topics II, New York, Academic Press.

Mardia, K. (1972). The Statistics of Directional Data. Academic Press, New York.

Muir, J. (1933). Treatise on the Theory of Determinants. Longmans, Green and Co.

Nevai, P. G. (1979). Orthogonal Polynomials. Memoirs of the Am. Math. Soc., No. 213.

Perron, O. (1913). Die Lehre von den Kettenbrüche. Leipzig and Berlin.

Polyá, G. and Szegö, G. (1972). Problems and Theorems in Analysis. Springer-Verlag.

Pukelsheim, F. (1980). On linear regression designs which maximize information. J. Stat. Planning and Inference, 4, 339-364.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications. John Wiley & Sons.

Shilov, G. Ye. (1965). Mathematical Analysis. Pergamon Press.

Sibson, R. (1974). D_A-optimality and duality. Progress in Statistics. Colloq. Math. Soc. Janos Bolya, 9, 677-692.

Silvey, S. D. and Titterington, D. M. (1974). A Lagrangian approach to optimal design. Biometrika, 61, 299-302.

Skibinsky, M. (1967). The range of the (n+1)th moment for distributions on [0,1]. J. Applied Probability, 4, 543-552.

Skibinsky, M. (1968). Extreme nth moment for distributions on [0,1] and the inverse of a moment space map. J. Applied Probability, 5, 693-701.

Skibinsky, M. (1969). Some striking properties of Binomial and Beta moments. Ann. Math. Statist., 40, 1753-1764.

Stigler, S. (1971). Optimal experimental designs for polynomial regression. JASA, 66, 311-318.

Studden, W. J. (1968). Optimal designs on Tchebycheff points. Ann. Math. Statist., 39, 1435-1447.

Studden, W. J. (1971). Optimal designs and spline regression. Optimizing methods in Statistics. Academic Press.

Studden, W. J. (1977). Optimal designs for integrated variance in polynomial regression. Statistical Decision Theory and Related Topics II, Academic Press.

Studden, W. J. (1978). Designs for large degree polynomial regression. Commun. Statist.-Theor. Meth. A7, 1391-1397.

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

Studden, W. J. (1981). Optimal designs for weighted polynomial regression using canonical moments. Statistical Decision Theory and Related Topics III, Academic Press.

Studden, W. J. (1982). Some robust-type D-optimal designs in polynomial regression. JASA, 77, 916-921.

Szegö, G. (1975). Orthogonal Polynomials, 4th ed., Am. Math. Soc. Coll. Pub., Vol. 23, New York Providence, Rhode Island.

Wall, M. S. (1948). Analytic Theory of Continued Fractions. New York: Van Nostrand.

Wu, C. F. (1978). Some iterative procedures for generating nonsingular optimal designs. Commun. Statist. A7, 14, 1399-1412.

Wynn, H. P. (1970). The sequential generation of D-optimal experimental designs. J. Roy. Statist. Soc. B, 34, 133-147.

Wynn, H. P. (1972). Results in the theory and construction of D-optimum experimental designs. J. Roy. Statist. Soc. Ser. B, 34, 133-147.