Saturated Designs For Multivariate Cubic Regression

by

W. Notz[1]
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #79-19

[1] Research supported by National Science Foundation Grant.

ABSTRACT

This paper deals with the problem of finding saturated designs for multivariate cubic regression on a cube which are nearly D-optimal. A finite class of designs is presented for the k dimensional cube having the property that the sequence of the best designs in this class for each k is asymptotically efficient as k increases. A method for constructing good designs in this class is discussed and the construction is carried out for $1 \leq k \leq 8$. These numerical results are presented in the last section of the paper.

Key Words: Cubic Regression, saturated designs, D-optimality, D-efficient, balanced arrays.

AMS Subject Classifications: Primary 62K05, 62J05; Secondary 05B30.

1. INTRODUCTION.

Let $I = [-1, +1]$ and let I^k be the k-fold Cartesian product of the closed interval I. Suppose on the basis of n observations of k-variate predictor variables $\underline{x}(m) = (x_1(m), x_2(m), \ldots, x_k(m))'$ in I^k (primes on vectors or matrices denote transposes), where m = 1, 2, ..., n, we wish to fit by least squares a third order model

$$(1.1)\quad y(\underline{x}(m)) = b_o + \sum_{i=1}^{k} b_i\, x_i(m) + \sum_{i=j}^{k} \sum_{j=1}^{k} b_{ij} x_i(m) x_j(m) + \sum_{i=j}^{k} \sum_{j=\ell}^{k} \sum_{\ell=1}^{k} b_{ij\ell} x_i(m) x_j(m) x_\ell(m) + e_m.$$

Here the b_i, b_{ij}, and $b_{ij\ell}$ are unknown real-valued parameters that we wish to estimate and the e_m are uncorrelated random variables with mean 0 and finite variance σ^2. The b_i, b_{ij}, and $b_{ij\ell}$ will be estimated by their least squares estimates. The purpose of this paper is to investigate how to choose the n points $\underline{x}(1), \ldots, \underline{x}(n)$ so as to minimize, in some sense, the covariance matrix of the least squares estimates.

Let $q = \binom{k+3}{3}$. From (1.1) it is easy to check that there are q parameters to be estimated and it is well known that in order for all the parameters to be estimable we need at least q observations.

An exact experimental design is a collection of points (repetition is possible) at which we shall take observations, one at each point in the collection. Any such collection of n points can be represented by a probability measure ξ, where $\xi(\underline{x}) = j(\underline{x})/n$ with $j(\underline{x})$ = number of times the point $\underline{x}$ appears in the collection. We shall denote experimental designs by such probability measures in what follows. The set of points $\underline{x}$ for which $\xi(\underline{x}) \neq 0$ is called the support of ξ.

Suppose we have an exact experimental design on the n points $\underline{x}(1), \ldots, \underline{x}(n)$. Let ξ denote the corresponding probability measure (ξ is often called the experimental design). Let

(1.2) $$f(\underline{x}) = (1, x_1, \ldots, x_k, x_1x_2, x_1x_3, \ldots, x_{k-1}x_k, x_1x_2x_3, x_1x_2x_4, \ldots, x_{k-2}x_{k-1}x_k, \ldots, x_1^2, x_1^2x_2, \ldots, x_1^2x_k, x_2^2x_1, x_2^2, x_2^2x_3, \ldots, x_2^2x_k, \ldots, x_k^2x_1, x_k^2x_2, \ldots, x_k^2, x_1^3, \ldots, x_k^3)'$$

where $\underline{x} = (x_1, \ldots, x_k)'$. $f(\underline{x})$ is a q x 1 vector and if $\underset{\sim}{b}$ is the corresponding vector of the coefficients b_i, b_{ij}, and $b_{ij\ell}$ in (1.1), we can rewrite (1.1) as

(1.3) $$y(\underline{x}(m)) = f'(\underline{x}(m))\,\underline{b} + e_m, \quad m=1, \ldots, n.$$

Now let $X(\xi)$ be the q x n matrix whose m-th column is $f(\underline{x}(m))$. If $\hat{\underline{b}}$ is the least squares estimate of $\underline{b}$, one can show $\hat{\underline{b}}$ has covariance matrix

(1.4) $$\text{cov}(\hat{\underline{b}}) = \sigma^2(X(\xi)X'(\xi))^{-1}$$

assuming $X(\xi)X'(\xi)$ is nonsingular (which will be the case in this paper). Recall our objective is to find a design ξ which minimizes $\text{cov}(\hat{\underline{b}})$ in some sense. We shall seek designs ξ which minimize the generalized variance, which is equivalent to minimizing $\det(X(\xi)X'(\xi))^{-1}$ or maximizing $\det X(\xi)X'(\xi)$. Such designs are called D-optimal designs.

It may be that the design ξ which maximizes $\det X(\xi)X'(\xi)$ takes a large number of observations which could be expensive or unrealistic to implement in practice. To avoid this difficulty we would like to restrict ourselves to designs which don't take too many observations.

For purposes of this paper we shall restrict ourselves to designs which take the minimum number of observations, q. Such designs are sometimes called saturated designs. If ξ is a saturated design then $X(\xi)$ is a q x q matrix and so $\det X(\xi)X'(\xi) = \det^2 X(\xi)$. Our objective then can be restated as seeking the saturated design ξ which maximizes $\det^2 X(\xi)$.

The solution of this problem in general is not known. The solution when k = 1 is well known (see Hoel (1958) or Guest (1958)) and the solution when k = 2 has been found by Dubova and Federov (1972). For k > 2 the saturated design ξ which maximizes $\det^2 X(\xi)$ is not known. In this paper we present a finite class of saturated designs for each k such that the best design in the class, although not necessarily optimal, has a certain optimality property and is in general "pretty good". The optimality property these designs have is an asymptotic one. If ξ_k denotes the best designs in the class presented in this paper for dimension k and ψ_k denotes the optimal saturated design for dimension k, then $\lim_{k \to \infty} [\det^2 X(\xi_k)/\det^2 X(\psi_k)]^{1/q} = 1$. This says that the sequence of best designs in this class for each k is asymptotically D-efficient. This follows easily from a result in chapter 4 of Notz (1978) and so no proof will be given here. In essence, this means that for large k the designs constructed by the method presented here are "pretty good". We shall carry out the construction of saturated designs using the method given in this paper up to k = 8 and we shall see that in the cases k = 1 and k = 2 where optimal designs are known, the designs we get aren't too bad.

2. METHOD OF CONSTRUCTION.

The method of construction given here is similar to that given for quadratic regression in Notz (1978).

Recall that for a saturated design , the columns of $X(\xi)$ are of the form $f(\underline{x})$, where $f(\underline{x})$ is given in (1.2). We are seeking to make $\det^2 X(\xi)$ large. Suppose we subtract the first row of $X(\xi)$ (a row all of whose entries are 1) from each row whose entries are of the form x_i^2. Also suppose we subtract the i + 1st row of $X(\xi)$ (a row all of whose entries are of the form x_i) from all rows whose entries are of the $x_i x_j^2$, $j = 1, 2, \ldots, k$. We do this for all $i = 1, \ldots, k$ and call the new matrix so formed $Z(\xi)$. These row operations do not change the value of $\det^2 X(\xi)$ so we have $\det^2 Z(\xi) = \det^2 X(\xi)$.

Let $c = 1 + k + \binom{k}{2} + \binom{k}{3}$. Consider the finite set of saturated designs Ξ where

(2.1) $\Xi(k) = \{\xi$; the support of ξ is $\underline{x}(1), \ldots, \underline{x}(q)$ where $\underline{x}(1), \ldots, \underline{x}(c)$ have all coordinates of the form ± 1, $\underline{x}(c + 1), \ldots, \underline{x}(c + k)$ have all first coordinates 0 but all other coordinates ± 1, $\underline{x}(c + k + 1), \ldots, \underline{x}(c + 2k)$ have all second coordinates 0 but all other coordinates $\pm 1, \ldots,$ $\underline{x}(c + k^2 - k + 1), \ldots, \underline{x}(c + k^2)$ have all k-th coordinates 0 but all other coordinates ± 1, and $\underline{x}(c + k^2 + 1), \ldots,$ $\underline{x}(c + k^2 + k) = \underline{x}(q)$ **have all the coordinates** $\pm {}^1/\sqrt{3}\}$

For example when k = 2 we have c = 4, q = 10 and one can check that a non-singular element of $\Xi(z)$ has $\underline{x}(1) = (1, 1)'$, $\underline{x}(2) = (1, -1)'$,

$\underline{x}(3) = (-1, 1)'$, $\underline{x}(4) = (-1, -1)'$, $\underline{x}(5) = (0, 1)'$, $\underline{x}(6) = (0, -1)'$, $\underline{x}(7) = (1, 0)'$, $\underline{x}(8) = (-1, 0)'$, $\underline{x}(9) = (1/\sqrt{3}, 1/\sqrt{3})'$, $\underline{x}(10) = (1/\sqrt{3}, -1/\sqrt{3})'$ as its support.

For any $\xi \in \Xi(k)$, $Z(\xi)$ takes on a particularly nice form. It has blocks $Z_1(\xi)$, $Z_2(\xi)$, ..., $Z_{k+2}(\xi)$ running down the diagonal and all entries below these blocks are 0. $Z_1(\xi)$ is a c x c matrix whose columns are of the form

(2.2) $g_1(\underline{x}) = (1, x_1, \ldots, x_k, x_1x_2, \ldots, x_{k-1}x_k, x_1x_2x_3, \ldots, x_{k-2}x_{k-1}x_k)'$.

$Z_i(\xi)$ for i = 2, ..., k + 1 is a k x k matrix whose columns are of the form

(2.3) $g_i(\underline{x}) = (x_1(x_{i-1}^2 -1), x_2(x_{i-1}^2 -1), \ldots, x_{i-2}(x_{i-1}^2 -1), x_{i-1}^2, x_i(x_{i-1}^2 -1), \ldots, x_k(x_{i-1}^2 -1))'$.

$Z_{1_{k+2}}(\xi)$ is a k x k matrix whose columns are of the form

(2.4) $g_{k+2}(\underline{x}) = (x_1^3 - x_1, x_2^3 - x_2, \ldots, x_k^3 - x_k)'$.

For any $\xi \in \Xi(k)$ it therefore follows that

(2.5) $\det^2 X(\xi) = \det^2 Z(\xi) = \prod_{i=1}^{k+2} \det^2 Z_i(\xi)$

since $Z(\xi)$ has the blocks $Z_i(\xi)$ down its diagonal and all entries 0 below these blocks.

It is not hard to see (use lemma 2.2 in chapter 3 of Notz (1978)) that there is a design ξ^* in $\Xi(k)$ that will simultaneously maximize all the $\det^2 Z_i(\xi)$. This will be the optimal design in $\Xi(k)$ (although not necessarily the optimal saturated design) and it is this design that we desire to construct.

The construction of ξ^* should proceed as follows. Choose the support $\underline{x}(1)$, ..., $\underline{x}(q)$ of ξ^* so that $\underline{x}(1)$, ..., $\underline{x}(c)$ make $\det^2 Z_1(\xi)$ as large as possible, so that for $i = 2, \ldots, k + 1$ that $\underline{x}(c + (i-2)k + 1), \ldots, \underline{x}(c + (i-1)k)$ make $\det^2 Z_i(\xi)$ as large as possible, and so that $\underline{x}(c + k^2 + 1), \ldots, \underline{x}(q)$ make $\det^2 Z_{k+2}(\xi)$ as large as possible. Finding such a ξ^* is a difficult combinatorial problem in general and the author can not give a general solution. However using balanced arrays and known ± 1 $k \times k$ matrices having maximal determinant one can construct a design in $\Xi(k)$ which is fairly good. We shall now examine briefly how to do this.

To find points $\underline{x}(1), \ldots, \underline{x}(c)$ in the support of a design ξ in $\Xi(k)$ so that $\det^2 Z_1(\xi)$ is large, we use balanced arrays of strength 6 which we now define. A $(-1, +1)$ matrix T of size $m \times N$ is called a balanced array of strength 6, size N, m constraints, and index set $\{\mu_0, \mu_1, \ldots, \mu_6$ if for every $6 \times N$ submatrix T_0 of T, every $(-1, +1)$ vector having exactly j entries which are $+1$ occurs exactly μ_j times $(j = 0, ., \ldots, 6)$ as a column of T_0. For $k \geq 6$ we shall choose the first c points in the support of our design ξ so that they form the columns of a balanced array of strength 6 of size $k \times q$ and so that $Z_1(\xi)$ has a large value of the square of its determinant. For a discussion of balanced arrays of strength 6, their construction, and formulas that can be used to compute the value of $\det^2 Z_1(\xi)$ see Shirakura (1976). For $k = 5$ one must use trial and error to determine the first c points in the support of ξ. For $k \leq 4$ all choices of the first c points in the support of ξ yielding a non-singular $Z_1(\xi)$ all give the same value of $\det^2 Z_1(\xi)$ (this is easy to verify). The results of Shirakura and trial and error are then used to find what appear to be reasonable choices for the first c points in the support of a good design.

To determine the remaining points in the support of our design $\xi \in \Xi(k)$ we need to know what is the k x k (-1, +1) matrix B yielding the maximum (or nearly maximum) value of $\det^2 B$. Values of B for various values of k and related discussion can be found in Ehlich (1964 a), Ehlich (1964 b), Yang (1966), Yang (1968), and section 17.4 and 17.6 of Raghavarao (1971), for example. Once B is known, the last q - c points in the support of $\xi \in \Xi(k)$ are chosen as follows. For i = 2, ..., k + 1 we choose the points $\underline{x}(c + (i-2)k + 1)$, ..., $\underline{x}(c + (i-1)k)$ in the support of ξ so that $Z_i(\xi) = B$. It is straightforward to verify that this can be done (recall the form of $Z_i(\xi)$ and what form $\underline{X}(c + (i-2)K + 1)$, ..., $\underline{x}(c + (i-1)k)$ must have so that $\xi \in \Xi(k)$). Finally we choose the last k points in the support of ξ so that $(3\sqrt{3}/2)\, Z_{k+2}(\xi) = B$. Again it is straightforward to check that this can be done.

This process yields as reasonable $\xi \in \Xi(k)$ and one gets for such a ξ

(2.6) $$\det^2 X(\xi) = (2/3\sqrt{3})^{2k} \det^2 Z_1(\xi)\ [\det B]^{2(k+1)}.$$

We have carried out this construction for k = 1, ..., 8 and our results shall be presented in the next section.

3. NUMERICAL RESULTS.

Although we have discussed finding a design so as to make $X(\xi)X'(\xi)$ large, the usual procedure is to find ξ so as to make the information matrix per unit variance $M(\xi)$ large. If ξ is a design having n points in its support then

(3.1) $$M(\xi) = \frac{1}{n} X(\xi)X'(\xi)$$

For fixed n, maximizing (in some sense) $M(\xi)$ and $X(\xi)X'(\xi)$ is equivalent.

One advantage of using $M(\xi)$ is that it normalizes the covariance matrix of our least squares estimates by the number of observations taken and hence allows one to compare designs taking different numbers of observations. Since we restricted ourselves to taking only $n = q$ observations the problems of maximizing $\det^2 X(\xi)$ and $\det M(\xi)$ are equivalent. However for purposes of standardization we shall present the values of $\det M(\xi)$ rather than $\det^2 X(\xi)$ for the designs given in this section. Since $\det M(\xi) = (\frac{1}{q})^q \det^2 X(\xi)$ it is easy to get $\det^2 X(\xi)$ from the value of $\det M(\xi)$.

Using the methods outlined in section 2 we get the follosing designs.

$k = 1$: Support is 1, -1, 0, $1/\sqrt{3}$

$k = 2$: Support is $(1, 1)'$, $(1, -1)'$, $(-1, 1)'$, $(-1, -1)'$, $(0, 1)'$, $(0, -1)'$, $(1, 0)'$, $(-1, 0)'$, $(1/\sqrt{3}, 1/\sqrt{3})'$, $(1/\sqrt{3}, -1/\sqrt{3})'$.

$k = 3$: Support is $(1, 1, 1)'$, $(1, 1, -1)'$, $(1, -1, 1,)'$, $(-1, 1, 1,)'$, $(1, -1, -1)'$, $(-1, 1, -1)'$, $(-1, -1, 1)'$, $(-1, -1, -1)'$, $(0, 1, 1)'$, $(0, 1, -1)'$, $(0, -1, 1)'$, $(1, 0, 1)'$, $(1, 0, -1)'$, $(-1, 0, 1)'$, $(1, 1, 0)'$, $(1, -1, 0)'$, $(-1, 1, 0)'$, $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})'$, $(1/\sqrt{3}. 1/\sqrt{3}, -1/\sqrt{3})'$, $(1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3})'$.

For $k = 4, 5, 6, 7, 8$ it becomes quite tedious to list all the points in the support of our designs. We therefore just indicate how to choose the first c points in the support of our design and list the columns of B.

k = 4: Choose as the first c = 15 points in the support the following (+1, -1) 4 x 1 vectors. The vector consisting of all -1's, the 4 vectors consisting of three -1's and one +1, the 6 vectors consisting of two -1's and two +1's, and the 4 vectors consisting of one -1 and three +1's. The columns of B are (1, 1, 1, 1)', (1, 1, -1, -1)', (1, -1, 1, -1)', and (1, -1, -1, 1)'.

k = 5: Choose as the first c = 26 points the following (+1, -1) 5 x 1 vectors. The vector consisting of all +1's, the 5 vectors having only one +1 coordinate, the 10 vectors having only two +1 coordinates, and the 10 vectors having only three +1 coordinates. The columns of B are (1, 1, 1, 1, 1)', (1, 1, -1, -1, -1)', (1, -1, 1, -1, -1)', (1, -1, -1, 1, -1)', (1, -1, -1, -1, 1,)'.

k = 6: Choose as the first c = 42 points the (+1, -1) 6 x 1 vectors that form the columns of the balanced array of strength 6, size 42, and 6 constraints with index set {1, 0, 1, 1, 0, 1, 0} given in Shirakura (1976). The columns of B are (1, 1, 1, 1, 1, 1)', (1, 1, 1, -1, -1, 1,)', (1, 1, 1, -1, 1, -1)', (1, -1, -1, -1, 1, 1)', (1, -1, 1, 1, -1, -1)', (1, 1, -1, 1, -1, -1)'.

$k = 7$: Choose the first $c = 64$ points the $(+1, -1)$ 7×1 vectors that form the columns of the balanced array of strength 6, size 64, and 7 constraints with index set $\{1, 1, 1, 1, 1, 1, 1\}$ (this is an orthogonal array) given in Shirakura (1976). The columns of B are (1, 1, 1, 1, 1, 1, 1,)', (1, 1, -1, 1, -1, -1, -1)', (1, -1, -1, 1, 1, 1, 1)', (1, -1, 1, 1, -1, -1, 1,)', (1, 1, 1, -1, 1, -1, 1,)', (1, 1, 1, -1, -1, 1, 1,)', (1, -1, 1, 1, 1, 1, -1,)'.

$k = 8$: Choose as the first $c = 93$ points the $(+1, -1)$ 8×1 vectors that form the columns of the balanced array of strength 6, size 64, and 8 constraints with index set $\{2, 2, 2, 1, 1, 2, 2\}$ given in Shirakura (1976). The columns of B are (1, 1, 1, 1, 1, 1, 1, 1)', (1, 1, -1, -1, 1, 1, -1, -1)', (1, -1, 1, -1, 1, -1, 1, -1)', (1, -1, -1, 1, 1, -1, -1, 1,)', (1, 1, 1, 1, -1, -1, -1, -1)', (1, 1, -1, -1, -1, -1, 1, 1,)', (1, -1, 1, -1, -1, 1, -1, 1)', (1, -1, -1, 1, -1, 1, 1, -1)'.

Next we list the values of $\det M(\xi)$ and the normalized value $[\det M(\xi)]^{1/q}$ for the designs given above. In addition we give the known maximum values of $[\det M(\xi)]^{1/q}$ over all saturated designs for $k = 1$ (see Hoel (1958) or Guest (1958)) and for $k = 2$ (see Dubova and Fedorov (1972)). For $k \geq 3$ it is not known what are the D-optimal saturated designs for cubic regression.

k	q	$\det M(\xi)$	$[\det M(\xi)]^{1/q}$	$\max[\det M(\xi)]^{1/q}$
1	4	2.315×10^{-3}	.2193	.2675
2	10	3.60×10^{-8}	.1801	.1896
3	20	3.41×10^{-17}	.1502	-
4	35	3.46×10^{-29}	.1538	-
5	56	2.80×10^{-48}	.1415	-
6	84	3.47×10^{-71}	.1450	-
7	120	4.34×10^{-99}	.1515	-
8	165	1.20×10^{-128}	.1680	-

In the cases $k = 1$ and $k = 2$ where better saturated designs are known, they should be used. However for $k \geq 3$ it is not known what are good designs. It is the author's opinion that the designs presented here have values of $[\det M(\xi)]^{1/q}$ close to the maximum for a saturated design.

4. ACKNOWLEDGEMENTS.

The author wishes to express his thanks to Professor Jack Kiefer who originally suggested this problem as a possible research topic while the author was a student at Cornell University. The author also wishes to thank Professor Bill Studden for his advice and helpful comments.

REFERENCES

1. Dubova, I.S. and Federov, V.V. (1972). Tables of Optimum Designs II (saturated D-optimal designs on a cube). Preprint No. 40, in Russian. Issued by Interfaculty Laboratory of Statistical Methods, Moscow University.

2. Ehlich, H. (1964 a). Determinantenabschätzungen für binäre Matrizen. *Math. Z.* 83, 123-132.

3. Ehlich, H. (1964 b). Determinantenabschätzungen für binäre Matrizen mit $n \equiv 3$ mod 4. *Math. Z.* 84, 428-447.

4. Hoel, P.G. (1958). Efficiency Problems in Polynomial Estimation. *Ann. Math. Statist.* 29, 1134-1145.

5. Guest, P.G. (1958). The Spacing of Observations in Polynomial Regression. *Ann. Math. Statist.* 29, 294-299.

6. Notz, W.I. (1978). Two Problems In Optimal Design Theory. Ph.D. Thesis, Cornell University.

7. Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments*. John Wiley and Sons, Inc., New York.

8. Shirakura, T. (1976). Optimal Balanced Fractional 2^m Factorial Designs of Resolution VII, $6 \leq m \leq 8$. *Ann. Statist.* 4, 515-531.

9. Yang, C.H. (1966). Some Designs for Maximal (+1, -1) Determinant of Order $n \equiv 2$ (mod 4). *Math. Comp.* 20, 147-148.

10. Yang, C.H. (1968). On the Designs of Maximal (+1, -1) Matrices of Order $n \equiv 2$ (mod 4). *Math. Comp.* 22, 174-180.