On Model Robust Design for Polynomial Regression

by

R.D. Cook and L.A. Thibodeau

Technical Report #281

Department of Applied Statistics
University of Minnesota
Saint Paul, Minnesota

February, 1977

ABSTRACT

Optimal designs ξ_n for polynomial regression of degree n are considered. A measure of model robustness is defined for general model spaces. The model robustness of ξ_n for polynomial model spaces is examined. Some robustness properties of the limiting design ξ_0 are considered for the limiting polynomial model space.

1. Introduction. Let $\underline{f}' = (f_0, f_1, \ldots, f_n)$ be a vector in R^{n+1} defined by the set of linearly independent functions f_i, $i = 0, 1, \ldots, n$, on some compact space χ. An experiment consists of selecting an x in χ and observing a random variable Y(x) with regression function $E(Y|x) = \underline{f}'\beta$ and variance σ^2. We assume that the functions, f_i, are known while the parameter vector $\beta' = (\beta_0, \beta_1, \ldots, \beta_n)$ and σ^2 are unknown. If ξ is a probability measure on χ then ξ defines an experimental design.

Exact designs concentrate mass $\xi(x_i)$ at points x_i, $i = 1,2,\ldots, r$, subject to the restriction that $N\xi(x_i) = n_i$ be an integer for all i. An exact design specifies that the experimenter is to take N uncorrelated observations n_i at x_i, $i = 1,2,\ldots, r$. The resulting covariance matrix of the least squares estimate of β is of the form

$$(\sigma^2/N)\; M^{-1}(\xi)$$

where the information matrix, $M(\xi)$, of the design has elements,

$$m_{ij}(\xi) = \int_\chi f_i(x)\; f_j(x)\; d\xi(x) \quad .$$

Approximate designs are not constrained by the requirement that $N\xi(x_i)$ be integral for all i. Here we consider only approximate designs. For justification see Fedorov (1972).

Many criteria have been proposed for optimizing the selection of a design, ξ. Generally, they all specify the selection of a design which minimizes some functional of the information matrix, $M(\xi)$. We consider the functionals,

(i) determinant $[M^{-1}(\xi)]$

(ii) $\sup_{x\in\chi} d_{\underline{f}}(x,\xi)$

where $d_{\underline{f}}(x,\xi) = \underline{f}'(x)\; M^{-1}(\xi)\; \underline{f}(x)$.

Designs minimizing these functionals are called D- and G-optimal designs, respectively. It is well known that D-optimal designs minimize the content of confidence ellipsoids for β while G-optimal designs minimize the maximum prediction variance over χ. Kiefer and Wolfowitz (1960) showed that for approximate designs D and G optimality are equivalent. Additional functionals and resulting equivalences are given in Fedorov (1972) and Kiefer (1974).

Clearly the G-optimality (D-optimaltiy) of a design depends on the model specification $\underline{f}$. In the next section we define model robust designs and consider the robustness of some designs for polynomial regression. Two conjectures are presented on the G-efficiencies of a class of designs for polynomial models.

The final two sections of this note are concerned with large degree polynomial regression: Let ξ_n be the G-optimal (D-optimal) design for n^{th} order polynomial regression and consider the sequence $\xi_1, \xi_2, \ldots$. Kiefer and Studden (1976) find the limiting design ξ_0 ($\xi_n \to \xi_0$ as $n \to \infty$) and examine some of its properties. Some of their findings are reviewed in Section 3. In Section 4 we present additional characteristics of ξ_0 and relate them to the problem of determining a model robust design for polynomial model spaces.

2. Model Robustness. The problem of determining a model robust design has been formulated in many settings (see Box and Draper (1959), Atkinson and Cox (1974), Huber (1974), Stigler (1971)). Let $\underline{F}$ be a space of models each specified by a function $f(x)$ defined over χ. We define a design ξ^* to be a model robust design with respect to $\underline{F}$ if ξ^* maximizes the minimum efficiency over $\underline{F}$. Since the D-efficiency of a design is no less than its G-efficiency (Atwood (1969)) we shall consider only the latter. The selection of a model

robust design by this or any of the other criteria that have been put forward is often extremely complicated. Thus, designs which are "almost robust" (i.e. have high efficiencies over $\underline{F}$) are of great interest. We find the following generalized definition of G-efficiency useful:

Definition 2.1. For fixed $\underline{f}$, the G-efficiency of the design ξ relative to the design ξ^* over S is

$$G_f(\xi,\xi^*; S) \equiv \inf_{x\in S} \sup_{x^*\in S} \frac{d_f(x^*,\xi^*)}{d_f(x,\xi)} .$$

This definition allows a comparison of the behavior of two arbitrary designs on some space S. When ξ^* is the G-optimal design for $\underline{f}$ and $S = \mathcal{X}$, this reduces to the usual definition of the G-efficiency (see Atwood (1969)) of the design ξ relative to the G-optimal design, ξ^*, over the full design space $\mathcal{X}$. Thus a model robust design, ξ^{**}, is defined by the property

$$\sup_\xi \inf_{f\in\underline{F}} G_f(\xi,\xi_f; \mathcal{X}) = \inf_{f\in\underline{F}} G_f(\xi^{**},\xi_f; \mathcal{X})$$

where ξ_f is the G-optimal design for the model $\underline{f}$.

In the following we consider model robustness for polynomial models;

$\mathcal{X} = [-1, 1]$,

$$\underline{f}_n'(x) = (1,x,x^2,\ldots, x^n) .$$

Let $\underline{F}_k = \{\underline{f}_n\}$, $n = 1,2,\ldots, k$

$$M_n(\xi) = \int_{-1}^{1} \underline{f}_n(x)\underline{f}_n'(x)\, d\xi(x)$$

$$d_n(x,\xi) = \underline{f}_n'(x)M_n^{-1}\underline{f}_n(x)$$

and for notational convenience, $G_n(\cdot, \cdot; \cdot)$ be the G-efficiency for $\underline{f}_n$. $\underline{F}_k$ is the polynomial model space of polynomials of degree not greater than k. For $k = 2$ a model robust design, say ξ^2, has mass $(\sqrt{10} - 1)/6$ at +1,-1 and the remaining mass at 0. Clearly a model robust design over a two point space is a design that has maximum G-efficiency among all designs with equal G-efficiency. For ξ^2,

$$G_1(\xi^2,\xi_1;\chi) = G_2(\xi^2,\xi_2;\chi) = .838$$

while for ξ_2

$$G_1(\xi_2,\xi_1;\chi) = .80 \ .$$

However for $k \geq 3$ finding a model robust design directly is a tedious algebraic exercise which provides little insight into a general solution. While we will address the properties of model robust designs in detail in a subsequent paper some immediate simplification of the problem is possible: In general, for each $k \geq 1$ there exists a symmetric model robust design with no more than $k + 2$ points of support ($k + 1$ when k is even) which must include -1, 0, and +1. To prove symmetry it is sufficient to note that the space of all model robust designs for fixed k is convex and that $d_n(x,\xi)$ is convex on the design space for all $n = 1,2,\ldots,k$ (see for example Stigler 1971). Symmetry reduces the problem to considering measures on [0,1]. Since the model robustness of a design is determined by its first $2k + 2$ moments, $k + 1$ are zero by symmetry, the dimension of the support is determined by the solution to the classical moment problem (see Shohat and Tamarkin, p. 42, 1943). Finally the inclusion of the points -1, 0, and +1 follows from a simple scaling argument.

Kendall and Stuart (1968), p. 161, among others, recommend the use of ξ_k, the G-optimal design for the largest order polynomial, when designing for $\underline{F}_k$.

Clearly, ξ_k will not be model robust unless $k = 1$, although it may be "almost robust." Our consideration of this recommendation resulted in the following conjectures:

Conjecture 2.1. For all $j \leq k$

$$\sup_{x\in[-1,1]} d_j(x,\xi_k) = d_j(1,\xi_k) \; \{ = d_j(-1,\xi_k)\}$$

Conjecture 2.2. For all $j \leq k$

$$d_j(1,\xi_k) = (k+1) - \frac{(k-j)^2}{k} \; .$$

For $j = k$ the conjectures are obviously true. We have verified by computation that the conjectures are true for $k \leq 10$. The proof of these conjectures seems to depend on properties of partial sums of orthogonal polynomials defined with respect to ξ_k.

3. The limiting design, ξ_0. The G-optimal design, ξ_n, for $\underline{f}_n$ minimizes

$$\sup_{x\in X} \underline{f}_n'(x) \, M_n^{-1}(\xi) \, \underline{f}_n(x)$$

among all designs ξ on $[-1,1]$. It is well known (Guest (1958), Fedorov (1972)) that one such design ξ_n concentrates mass $1/(n+1)$ at the zeros of $(1-x^2)P_n'(x)$ where $P_n(x)$ is the n^{th} order Legendre polynomial. The following theorem is given in Fedorov (1972), page 91, and by Kiefer and Studden (1976).

Theorem 3.1. The sequence ξ_n, $n = 1,2,\ldots$, of G-optimal designs converges weakly to ξ_0 where ξ_0 has the density $[\pi(1 - x^2)^{\frac{1}{2}}]^{-1}$.

The proof consists of showing that if $x_1,\ldots, x_{n+1}$ is the support of ξ_n then $\theta_i = \cos^{-1}x_i$ are uniformly distributed in the interval $[0,\pi]$ as $n \to \infty$. Thus, $\xi_n \overset{w}{\to} \xi_0$ where ξ_0 is the distribution of $X = \cos\theta$ and θ is uniform on $[0,\pi]$.

Let

$$d_n(\xi) = \sup_{x\in X} d_n(x, \xi) \; .$$

It is well known that $d_n(\xi_n) = n + 1$ and that the supremum is obtained at the points of support of ξ_n.

The following additional results by Kiefer and Studden will be useful:

i) $d_n(x,\xi_0) = n + \frac{1}{2} + \frac{1}{2} U_{2n}(x)$

$$d_n(\cos\theta,\xi_0) = n + \tfrac{1}{2} + \frac{\sin[(2n+1)\theta]}{2\sin\theta}$$

where $U_k(x)$ are the Chebyshev polynomials of the second kind.

ii) $\sup_{x\in\chi} U_k(x) = U_k(1) \equiv \lim_{x\uparrow 1} U_k(x) = k + 1$

for all $k = 0, 1, 2, \ldots$

thus $d_n(\xi_0) = 2n + 1$.

iii) $$G(\xi_0,\xi_n; \chi) = \frac{n+1}{d_n(\xi_0)} = \frac{n+1}{2n+1} \to \tfrac{1}{2} \qquad \text{as } n \to \infty .$$

iv) For each $\varepsilon > 0$ there exists a design ξ_ε such that

$$\liminf_{n\to\infty} \frac{n+1}{d_n(\xi_\varepsilon)} \geq 1 - \varepsilon.$$

The first two results are necessary for the third which shows that the limit of the sequence of G-efficiencies is ½. This result is somewhat surprising and led Kiefer and Studden to question the existence of designs with limiting G-efficiency equal to one. The fourth result is the product of their inquiry. Note that property ii) proves Conjectures 2.1 and 2.2 for $k = \infty$.

4. Robustness of ξ_0. Consider the polynomial space $\underline{F}_k$ and suppose now that k is allowed to become arbitrarily large, then $\xi_k \to \xi_0$ and the G-efficiency of ξ_0 for n^{th} order regression becomes $(n+1)/(2n+1)$. Recall $\inf_{f\in \underline{F}_\infty} (n+1)/(2n+1) = \frac{1}{2}$. Clearly, we would prefer a substantially larger efficiency for a design to be adequate for $\underline{F}_k$. However, as will be shown, this result is a consequence of

a singularity in $d_n(x,\xi_0)$ at $\pm$ 1 as $n \to \infty$. This tends to give a distorted view of the robustness properties of ξ_0.

Let $S_\varepsilon = [-1+\varepsilon, 1-\varepsilon]$, $1 > \varepsilon > 0$. The following theorem shows the behavior at ξ_0 relative to ξ_n over S_ε as $n \to \infty$.

Theorem 4.1. For all $0 < \varepsilon < 1$

$$\lim_{n\to\infty} G_n(\xi_0,\xi_n;S_\varepsilon) = 1 .$$

Proof: Clearly, for any $0 < \varepsilon < 1$ there exists an n_ε such that for all $n \geq n_\varepsilon$

$$\sup_{x\varepsilon S_\varepsilon} d_n(x,\xi_n) = n + 1 .$$

Thus,

$$G_n(\xi_0, \xi_n;S_\varepsilon) = \inf_{x\varepsilon S_\varepsilon} \frac{n+1}{d_n(x,\xi_0)}$$

for $n \geq n_\varepsilon$. Let $x = \cos\theta$, $\theta\varepsilon(\cos^{-1}(1-\varepsilon),\cos^{-1}(-1+\varepsilon))$. From result i,

$$\frac{n+1}{d_n(\cos\theta,\xi_0)} = (1 + \frac{1}{n}) \left[1 + \frac{1}{2n} + \frac{\sin(2n+1)\theta}{2n(\sin\theta)}\right]^{-1} .$$

For all n and $\theta \varepsilon (0,\pi)$

$$0 < |\sin(2n+1)\theta| \leq 1 .$$

Thus,

$$\frac{\sin(2n+1)\theta}{2n(\sin\theta)} \to 0 \text{ as } n \to \infty \text{ for all } \theta \varepsilon [\cos^{-1}(1-\varepsilon),\cos^{-1}(-1+\varepsilon)].$$

The result follows.

The theorem shows that for any $1 > \varepsilon > 0$, the limiting design ξ_0 is asymptotically equivalent in the sense of G-efficiency to the G-optimal design over the interval S_ε. This result is consistent with a fundamental property of G-optimal designs in the finite case: The G-optimal design places measure one on the points of maximum prediction variance. Therefore, if a continuous design with support χ is G-optimal it seems reasonable to expect that the prediction variance must be maximized at each point in χ. The following lemma illustrates this.

Lemma 4.1. For all $x \varepsilon \chi$

$$\lim_{n \to \infty} \frac{n+1}{d_n(x,\xi_0)} = 1 .$$

Proof: For $x \varepsilon (0,1)$ the result is immediate from the proof of Theorem 4.1. The result follows (see ii, section 3).

Note that for $0 < \varepsilon < 1$,

$$\lim_{n\to\infty} \inf_{x\varepsilon S_\varepsilon} (n+1)/d_n(x,\xi_0) = \inf_{x\varepsilon S_\varepsilon} \lim_{n\to\infty} (n+1)/d_n(x,\xi_0) = 1 .$$

However, for $\varepsilon = 0$,

$$\tfrac{1}{2} = \lim_{n\to\infty} \inf_{x\varepsilon\chi} (n+1)/d_n(x_0,\xi_0) \neq \inf_{x\varepsilon\chi} \lim_{n\to\infty} (n+1)/d_n(x,\xi_0) = 1$$

The left hand side represents the limit of the sequence of G-efficiencies, while the right hand side may be interpreted as the G-efficiency for the limiting design ξ_0 and the limiting model f_0, $f_n \to f_0$ as $n \to \infty$.

The previous results suggest that if the experimenter is mainly interested in predicting on some subset of the open interval $(-1,1)$ and the degree of the polynomial is unknown then ξ_0 could be a reasonable (albeit not optimal) choice. In what follows we will present a brief outline of the model robust characteristics of ξ_0 on selected subsets of $(-1,1)$.

For any finite n the value achieved by $G_n(\xi_0,\xi_n;S_\varepsilon)$ is a function of ε and depends on the extrema of the Chebyshev polynomials, $U_{2n}(x)$, in the interval S_ε. As n increases the location and number of extrema change as well as the values of $U_{2n}(x)$ at the extrema. Some insight into the relationship between ε and $G_n(\xi_0,\xi_n;S_\varepsilon)$ can be obtained by considering the smallest value of ε that corresponds to the largest possible G-efficiency for each $\underline{f}_n$, n = 1, 2, ... Table 1 gives $\sup_{0<\varepsilon\leq\delta_n} G_n(\xi_0,\xi_n;S_\varepsilon)$ and the corresponding value of ε for $n \geq 2$, where δ_n is selected such that S_{δ_n} contains all the roots of $U_{2n}(x)$. Clearly if ε is allowed to become arbitrarily close to 1 then for n odd there exists an ε^* such that

$$G_n(\xi_0,\xi_n;S_{\varepsilon^*}) > 1$$

while for n even

$$G_n(\xi_0,\xi_n;S_\varepsilon) \leq 1$$

for all ε. Our restriction of ε to the interval $(0,\delta_n]$ thus excludes only that portion of χ at the boundaries $\pm$ 1. For n = 1

$$G_1(\xi_0,\xi_1;S_\varepsilon) = 2\left[2.5 + \frac{\sin 3\delta}{2\sin \delta}\right]^{-1}$$

where $\delta = \cos^{-1}(1 - \varepsilon)$.

Note that for all $\gamma > \varepsilon$ and all n

$$G_n(\xi_0,\xi_n;S_\varepsilon) \leq G_n(\xi_0,\xi_n;S_\gamma) .$$

The results in Table 1 do not contradict Theorem 4.1 since both ε and n were allowed to vary.

Table 2 gives $G_n(\xi_0,\xi_n;S_\varepsilon)$ for ε = .01, .05 and .1. Table 2 shows that the convergence of $G_n(\xi_0,\xi_n;S_\varepsilon)$ is not monotonic, however this characteristic can be easily explained by considering some fundamental properties of $d_n(x,\xi_0)$ or equivalently $U_{2n}(x)$. The G-efficiency $G_n(\xi_0,\xi_n;S_\varepsilon)$ depends only on the $\sup_{x\in S_\varepsilon} U_{2n}(x)$. For fixed ε, $U_{2n}(x)$ has fewer than n local maxima in S_ε. If $x_n^{(m)}$ corresponds to the m^{th} local maxima where $x_n^{(1)} < x_n^{(2)} < . . . < x_n^{(2n-1)}$, then for m fixed

$$|U_{2n}(x_n^{(m)})| < |U_{2(n+1)}(x_{n+1}^{(m)})|$$

and

$$|x_n^{(m)}| < |x_{n+1}^{(m)}| \leq 1 .$$

Also, as $n \to \infty$, $|x_n^{(m)}| \to 1$ and therefore the irregularity of $G_n(\xi_0, \xi_n : S_\varepsilon)$ is caused by the increasing values of the local maxima and the passage of $x_n^{(m)}$ corresponding to these maxima beyond the interval S_ε. For very large n the maxima are "tightly packed" in the interval [-1,1] thus the convergence is much smoother.

By presenting the efficiency of ξ_0 for S_ε we are not arguing that the variance in $[-1,1] - S_\varepsilon$ be ignored but rather that some discounting needs to be considered. Table 1 clearly shows that for large order polynomial regression ξ_0 has high efficiency relative to ξ_n for essentially all subsets of (-1,1). Further while recognizing that ξ_0 was not derived as a model robust design for $\underline{F}_\infty$ the efficiencies in Table 2 are probably not unreasonable for such a large model space. In a practical view it might be desirable if the efficiencies were decreasing for larger order polynomials and not increasing as they are for ξ_0 on S_ε. Such a characteristic might be used to select a specific design from the class of all model robust designs for some model space $\underline{F}_k$.

TABLE 1

Sup G-efficiency in S_ε for restricted ε;

$$\sup_{0<\varepsilon\leq\delta_n} G_n(\xi_0,\xi_n:S_\varepsilon) = G_n$$

n	2	3	4	5	6	7	8	9	10	11	12	50	100	1000
G_n	1.0	.99	.97	.96	.95	.94	.93	.93	.93	.92	.92	.90	.89	.89
	.134	.074	.046	.031	.022	.017	.013	.011	.009	.007	.006	$4x10^{-4}$	10^{-4}	10^{-6}

TABLE 2

G-efficiency of ξ_0 in S_ε for ε = .01, .05, .10 .

n	1	2	3	4	5	6	7	8	9	10	11	12	50	100	1000
ε															
.01	.68	.62	.62	.63	.66	.71	.76	.83	.90	.93	.92	.92	.94	.975	.997
.05	.71	.73	.84	.97	.96	.95	.94	.93	.93	.93	.92	.92	.98	.989	.998
.10	.76	.88	.99	.97	.96	.95	.94	.93	.96	.97	.97	.97	.99	.994	.999

REFERENCES

[1] Atkinson, A.C. and Cox, D.R. (1974). Experiments for model discrimination. J. Roy. Statist. Soc. Ser. B 36 321-334.

[2] Atwood, C.L. (1969). Optimal and efficient design of experiments. Univ. of Minnesota, School of Statistics, Tech. Rpt. 106.

[3] Box, G.E.P. and Draper, N.R. A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54 622-654.

[4] Fedorov, V.V. (1972) Theory of Optimal Experiments. Academic Press, New York.

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

[6] Huber, P.J. (1974). Robustness and designs. In A Survey of Statistical Designs in Linear Models, ed. J.N. Srivastava. North Holland, Amsterdam.

[7] Karlin, S. and Studden, W.J. (1966). Optimal experimental designs. Ann. Math. Statist. 37 783-815.

[8] Kendall, M.G. and Stuart, A. (1968). The Advanced Theory of Statistics, Vol. 3. Hafner Publishing Co., New York.

[9] Kiefer, J. and Studden, W.J. (1976). Optimal designs for large degree polynomial regression. Ann. Math. Statist. 4 1113-1123.

[10] Kiefer, J. and Wolfowitz, J. (1959). Optimum Designs in Regression Problems. Ann. Math. Statist. 30 271-292.

[11] Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canad. J. Math. 12 363-366.

[12] Kiefer, J. (1961). Optimum designs in regression problems, II. Ann. Math. Statist. 32 298-325.

[13] Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Math. Statist. 2 849-879.

[14] Kiefer, J. (1975). Optimal design: Variation in structure and performance under change of criterion. Biometrika 62 277-288.

[15] Shohat, J.H. and Tamarkin, J.D. (1943). The Problem of Moments, American Mathematical Society Surveys, No. 1, New York.

[16] Stigler, S.M. (1971). Optimal experimental design for polynomial regression. J. Amer. Statist. Assoc. 66 311-318.

[17] Szego, G. (1975). Orthogonal Polynomials. Amer. Math. Soc. Colloq. Publ., Vol. XXIII. Amer. Math. Soc., Providence, R.I.