Optimal Designs for Multivariate Polynomial Extrapolation*

by

W.J. Studden

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series #219

January 1970

*This research was supported by the National Science Foundation Contract GP-8986.

Optimal Designs for
Multivariate Polynomial Extrapolation*

by

W.J. Studden

Purdue University

Let $f = (f_1, f_2, \ldots, f_k)$ be a vector of linearly independent continuous functions on a compact set X in Euclidean m-space. For each "level" x in X an experiment can be performed whose outcome is a random variable $Y(x)$ with mean value $\sum_{i=1}^{k} \theta_i f_i(x)$ and variance σ^2, independent of x. The functions $f_1, f_2, \ldots, f_k$ are called the regression functions and assumed known to the experimenter, while the vector of parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ and σ^2 are unknown. We will be concerned here with the problem of estimating the regression function $\sum_{i=1}^{k} \theta_i f_i(\bar{x})$, at a point $\bar{x}$ outside of X, by means of a finite number of uncorrelated observations $Y(x_i)$. The design problem is one of selecting the levels x_i in X at which to experiment. The result here is approximate in that we consider a design to be a arbitrary probability measure on X. For a more complete discussion of the model see Kiefer (1959) or Karlin and Studden (1966).

For the case $X=[-1,1]$ and $\sum_{i=1}^{k} \theta_i f_i(x) = \sum_{i=1}^{k} \theta_i x^{i-1}$, Hoel and Levine (1964) showed that the optimum design for estimating $\sum_{i=1}^{k} \theta_i \bar{x}^{i-1}$ (for any $\bar{x} \notin [-1,1]$) was supported on the points $x_\nu = -\cos \frac{\nu\pi}{k-1}$ $\nu=0,1,\ldots,k-1$. Kiefer and Wolfowitz (1965), Studden (1968) and Studden and Karlin (1966) give further results for the case where the system $\{f_i(x)\}_1^k$ is a Tchebycheff system. Hoel (1965) gives

*This research was supported by the National Science Foundation Contract GP-8986.

a discussion of the extrapolation problem in multidimensions when the regression function is essentially of a product type.

In the present paper we consider the case where the regression function is a polynomial in m dimensions of degree less than or equal to n. The domain X will be a compact <u>convex</u> subset of the Euclidean m-space. Thus we take our f_i to be the functions $x_1^{\alpha_1} \ldots x_m^{\alpha_m}$ where the α_j are nonnegative integers and $\sum_{j=1}^{m} \alpha_j \leq n$. The number of such functions is $k = \binom{n+m}{m}$ and we assume that they are arranged in some fixed order.

<u>Optimal Design</u>. The optimal extrapolating design is described as follows. Consider a line through $\bar{x}$ which intersects the convex set X at two points, say a and b, such that the tangent hyperplanes at a and b are parallel. (The line in question exists but is not necessarily unique). The optimal design for extrapolating to $\bar{x}$ is now obtained by using the one dimensional result for polynomials of degree n on the line through a and b. Thus we consider the transformation $x(\alpha)=[(1-\alpha)a+(1+\alpha)b]/2$, such that $x(-1)=a$ and $x(+1)=b$. The optimal design concentrates on the points $x_\nu = x(\alpha_\nu)$ where $\alpha_\nu = -\cos \frac{\nu\pi}{n}$, $\nu=0,1,\ldots,n$. The optimal weights p_ν, $\nu=0,1,\ldots,n$ can be found as in the 1-dimensional case by solving the linear equations $\beta g(\bar{\alpha}) = \sum_{\nu=0}^{n} (-1)^\nu p_\nu g(\alpha_\nu)$ where $x(\bar{\alpha}) = \bar{x}$ and $g(\alpha)=(1,\alpha,\ldots,\alpha^n)$. The minimal variance is $\sigma^2 T_n^2(\bar{\alpha})/N$ where $T_n(\alpha)$ is the nth degree Tchebycheff polynomial of the 1st kind and N is the number of observation. The result is approximate in the sense that the numbers Np_ν, corresponding to the number of observations taken at x_ν may not be integer valued.

<u>Examples</u>. We shall not give any detailed numerical examples here. Instead we consider some simple discussion involving the convex set X. The existence of the line segment, on which the observations are taken, is shown intuitively in 2-dimensions as follows. We consider a ray emanating from $\bar{x}$ and let it sweep

through 360 degrees starting in a position not intersecting X. When the ray just touches X the two points of intersection a and b coincide. They then roll around the set X on opposite sides. The corresponding supporting hyperplanes must at some point be parallel.

In cases where the set X is symmetric about the origin the line segment in question is easily seen to go thru the origin. This is the case for example with the unit ball $X = \{x=(x_1,\dots,x_m) \mid \sum_1^m x_i^2 \le 1\}$. That the line segment and the optimal design are not unique is seen from the m-cube $X=\{x \mid \max|x_i| \le 1\}$. If m=2 and n=2 and $\bar{x} = (2,0)$ one can use the three points $(1,\rho)$ $(0,2\rho)$ $(-1,3\rho)$ for any ρ with $|\rho| \le 1/3$. It can easily be shown that any convex combination of optimal designs (viewed as measures) is again optimal. Thus one can produce optimal designs supported on any multiple of 3 points.

For the m-simplex, represented in m+1 coordinates as $X = \{x=(x_0,x_1,\dots,x_m) \mid \sum_0^m x_i = 1\}$, the line segment goes thru the "opposite" vertex.

It was originally thought that when drawing the line segment thru $\bar{x}$ and intersecting X at a and b, the required line was such that the distance from a to b was a maximum, thus extrapolating to $\bar{x}$ with the longest one dimensional set thru X. This is seen to be false by considering m=2 and taking X to be an extremely "elongated" ellipse.

Proof of optimality. The proof, as well as the result itself, was motivated by a paper by Rivlin and Shapiro (1961). We follow closely the proof given for the 1-dimensional case in Karlin and Studden (1966).

For a given design or probability measure ξ on X the variance of the best linear unbiased estimate of $\sum_{i=1}^{k} \theta_i f_i(\bar{x})$ is proportional to

$$V(\bar{x},\xi) = \sup_d \frac{(d,f(\bar{x}))^2}{\int (d,f(x))^2 \, d\xi(x)}$$

where $f(x) = (f_1(x),\ldots,f_k(x))$, d is a k-vector and $(d,f) = \sum_i d_i f_i$. The design ξ is said to be optimal for extrapolating to $\bar{x}$ if it minimizes $V(\bar{x},\xi)$.

We consider the line segment thru $\bar{x}$ cutting the convex set X at two points a and b so that the support planes are parallel. The existence of such a line segment is given in Rivlin and Shapiro (1961) for X strictly convex. The proof can be extended to just convexity. Now take a line segment thru $\bar{x}$ and perpendicular to the two parallel support planes. We consider a new orthogonal coordinate system with this line as the first coordinate or axis. The origin at the midpoint between the two support planes and the scale on this axis so that the distance from the new origin to either support plane is one. This involves a change of variable $Z = A(x-c)$ where c is the new origin and A is a nonsingular m×m matrix. If $B=A^{-1}$ then $x=Bz+c$ and due to the polynomial nature of our component functions we may write $f(Bz+c) = \Lambda f(z)$ where Λ is some nonsingular k×k matrix. In this case we can work in the z coordinate system with the same vector f(z) since (with the usual abuse of notation)

$$d(\bar{x},\xi) = \sup_d \frac{(d,\Lambda f(\bar{z}))^2}{(d,\Lambda f(z))^2 \, d\xi(Bz+c)}$$

$$= \sup_e \frac{(e,f(\bar{z}))^2}{\int (e,f(z))^2 \, d\eta(z)}$$

where $d\eta(z) = d\xi(Bz+c)$.

We use the geometrical result of Elfving which states that: If $R_+=\{f(z)|z\varepsilon X\}$ and $R_- = -R_+$ and R denotes the convex hull of $R_+ \cup R_-$ then the design η is optimal for $\bar{z}$ if there exists a function $\varphi(z) \pm 1$ such that

1. $\int f(z)\ \varphi(z)\ d\eta(z) = \beta f(\bar{z})$ for some scalar β and
2. $\beta f(\bar{z})$ is on the boundary of R.

Moreover $\beta^{-2} = \min\limits_{\eta} d(\bar{z},\eta)$.

To apply this result we rely heavily on Hoel's one-dimensional extrapolation result. Let z_1 denote the first component of z, $g(z_1) = (1,z_1,\ldots,z_1^n)$ and η_1 the optimal one-dimensional design for extrapolating to $\bar{z}_1$. Then the one-dimensional result states that

$$(1)\qquad \int g(z_1)\ \varphi_1(z_1)\ d\eta_1(z_1) = \frac{1}{|T_n(\bar{z}_1)|}\ g(\bar{z}_1)$$

where T_n is the nth degree Tchebycheff polynomial of the 1st kind. Moreover the coordinates of T_n define the support plane to R (for the g system) at the boundary point (1). (If d* denotes the vector of coordinates of T_n then the support plane is either $(d^*,y) = +1$ or $(d^*,y) = -1$).

The procedure now involves showing that the same result holds for the system f(z) where η_1 (which is presently supported on the z_1 axis) is replaced by a measure η_0 obtained by moving the mass of η_1 perpendicularly off of the z_1 axis to the line segment from a to b.

Let e* denote the k-vector with components corresponding to T_n whenever z_1 appears and zeros elsewhere. Then e* gives the support plane to R (in the f system) at the point $f(\bar{z})/|T_n(\bar{z}_1)|$. Thus it suffices to show that

(2) $$\int f(z)\,\varphi_0(z)\,d\eta_0(z) = \frac{1}{|T_n(\bar{z}_1)|}\,f(\bar{z})$$

where $\varphi_0(z) = \varphi_1(z_1)$ and φ_1 is given in (1). Note that componentwise equation (2) holds for any component involving only z_1 while the right hand side, for any component involving something other than z_1, is zero, since $\bar{z} = (\bar{z}_1, 0, \ldots, 0)$. It thus suffices to show that

(3) $$\int \varphi_0(z) \prod_{i=1}^{m} z_i^{\alpha_i}\,d\eta_0(z) = 0$$

whenever $\alpha_i \neq 0$ for some $i = 2, \ldots, m$.

Now the mass of η_1 was moved perpendicularly from points $z_1(\nu)$ on the z_1 axis to the line segment from a to b so that the mass of η_0 is now on points $\bar{z} + t_\nu(b-\bar{z}), \nu = 0,1,\ldots,n$ where $t_\nu = (z_1(\nu)-\bar{z}_1)/(b_1-\bar{z}_1)$. Omitting the factor $\prod_{i=2}^{m} \left(\frac{b_i}{b_i-\bar{z}_1}\right)^{\alpha_i}$ equation (3) can then be written as

(4) $$\int z_1^{\alpha_1} \prod_{i=2}^{m} (z_1-\bar{z}_1)^{\alpha_i}\,\varphi_0(z)\,d\eta_0(z)$$

$$= \int \sum_{\ell_2=0}^{\alpha_2} \cdots \sum_{\ell_m=0}^{\alpha_m} \binom{\alpha_2}{\ell_2} \cdots \binom{\alpha_m}{\ell_m} z_1^{\gamma}\,(-\bar{z}_1)^{\delta}\varphi_0(z)\,d\eta_0(z)$$

where $\gamma = \alpha_1 + \sum_2^m \ell_i$ and $\delta = \sum_2^m (\alpha_i - \ell_i)$. Now by equation (1)

$$\int z_1^{\gamma}\,\varphi_0(z)\,d\eta_0(z) = \frac{\bar{z}_1^{\gamma}}{|T_n(\bar{z}_1)|}\ .$$

Therefore, omitting the factor involving T_n, equation (4) becomes

$$\sum_{\ell_2=0}^{\alpha_2} \cdots \sum_{\ell_m=0}^{\alpha_m} \binom{\alpha_2}{\ell_2} \cdots \binom{\alpha_m}{\ell_m} (-1)^{\delta} \bar{z}_1^{\rho}$$

where $\rho = \sum_1^m \alpha_i$. Since $\delta = \sum_2^m (\alpha_i - \ell_i)$ this expression is zero by the binomial theorem.

Further Remarks. The optimal extrapolating design enables one to determine the support of further optimal designs. We will show that one can easily find the optimal design for estimating the coefficient of the term x_i^n for each i and also the sum of the coefficients of all of the nth degree terms.

If c denotes a k-vector, we consider estimating the linear function $(c,\theta) = \Sigma c_i \theta_i$. Suppose that the optimal designs for estimating the sequence of linear functions $(c^{(r)},\theta)$ have support on finite sets $B_r \subset X$. The number of points in B_r can always be assumed to be at most $k(k+1)/2 + 1$. If $c^{(r)}$ converges to c and the sets B_r converge to B(in the obvious manner) then there is an optimal design for estimating (c,θ) that is supported on B. This follows readily from the compactness of X, the continuity of f_i and Elfving's Theorem. This procedure was noted previously by Kiefer.

To obtain the optimal design for estimating x_1^n we take $\bar{x}(r) = (\bar{x}_1(r),0,\ldots,0)$ and $c^{(r)} = f(\bar{x}(r))/|\bar{x}_1(r)|^n$. Letting $\bar{x}_1(r) \to \infty$ the vector $c^{(r)}$ converges to a vector with a one in the appropriate coordinate and zeros elsewhere. To obtain the design for estimating the sum of the coefficients of all of the nth degree terms (including the "cross" terms) we let $\bar{x}(r) = (\bar{x}_1(r),\bar{x}_1(r),\ldots,\bar{x}_1(r))$ and take $c^{(r)}$ as before. In each of the above cases the appropriate weights at the $n+1$ "Tchebycheff points" are proportional to 1:2:2:...:2:1.

As an example consider $m=n=2$. If X is the unit circle then the design for the coefficient of x_1^2 is on (-1,0),(0,0) and (1,0) and x_2^2 is on (0,-1),(0,0)(0,1). The design for estimating the sum of the coefficients of x_1^2, x_2^2 and x_1x_2 is on the 45 degree line at the center and the two points on the circumference. Note further that if we take X to be any circle (not necessarily centered at the origin) then the designs remain the same in the sense that they are still on lines thru the center of the circle; the lines being the horizontal, vertical and the 45 degree lines again.

In the multivariate case we may wish to determine whether the coefficients of the nth degree terms are all zero, in which case the regression is of degree at most $n-1$. We note that the optimal design, for estimating the sum of the coefficients of the nth degree terms, is not readily effective for this purpose. This is due to the fact that the individual terms may be nonzero but cancel each other out to get a sum equal to zero.

References

1. Hoel, P.G. and Levine, A. (1964). Optimal spacing and weighting in polynomial regression. Ann. Math. Statist. 35, 1553-1560.

2. Hoel, P.G. (1965). Optimum desings for polynomial extrapolation. Ann. Math. Statist. 36, 1483-1493.

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

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

5. Kiefer, J. and Wolfowitz, J. (1965). On a theorem of Hoel and Levine on extrapolation designs. Ann. Math. Statist. 36, 1627-1655.

6. Rivlin, T.J. and Shapiro, H.S. (1961). A unified approach to certain problems of approximation and minimization. J. Soc. Indust. Appl. Math. Vol. 9, 670-699.

7. Studden, W.J. (1968). Optimal designs on Tchebycheff points. Ann. Math. Statist. 39, 1435-1447.