AN ALGORITHM FOR THE CHOICE OF OPTIMAL RESPONSE SURFACE DESIGNS

BY

M. SINGH, A. DEY AND R.K. MITRA

I.A.S.R.I., New Delhi

(Received : August, 1976)

1. INTRODUCTION AND PRELIMINARIES

Several optimality criteria in general regression problems have been discussed by Kiefer [2] and Kiefer and Wolfowitz [3]. Many designs are available in literature for fitting the second order response surfaces. In particular, central composite designs, Box-Behnken designs, uniform shell designs are recommended for practical purposes. Lucas [4] has tabulated the *G*-efficiencies and *D*-efficiencies of some designs of above types. The choice between two designs can be made on the basis of comparing the efficiencies of the two designs.

Lucas [4] made use of the computer for calculation of ultimate efficiencies of various designs which involves a long process. In the present paper we have evolved an algorithm for comparing various second order response surface designs.

Following Kiefer [2] a design can be treated as a measure ξ on χ by $\xi(x)$, representing the proportion of observations taken at a point x, χ being the experimental region usually taken as $[-1, 1]$. In an N point design with n_i observations at x_i $\left(\sum n_i=N\right)$, we have

$$\xi(x_j)=\begin{cases} 0 \text{ if there are no observations at } x_j \\ n_j/N \quad \text{if} \quad n_j>0 \end{cases}$$

In a discrete N point design, ξ takes value $1/N$ and defines an exact design.

Consider the usual regression set up given by

$$E(y_x)=\sum_{i=1}^{p} \beta_i f_i(x)=\underline{f}'(x)\,\underline{\beta}$$

$$V(y_x)=\sigma^2$$

where $\underline{\beta}=(\beta_1, \ldots\ldots, \beta_p)'$

is column vector of regression coefficients and σ^2 is the per observation variance. Given an integer N, it is desired to select a set

$$\underset{\sim}{x}=(\underset{\sim}{x}_1, \ldots\ldots, \underset{\sim}{x}_N)'$$

of N points in χ for which the random variables

$$Y=(y_{x_1}, \ldots\ldots, y_{x_N})'$$

are observed. We can also write

$$E(y_x)=\underline{f}'(x)\ \underline{\beta}$$

or

$$E(Y)=\underset{\sim}{X}\ \underline{\beta}$$

where i-th row of $\underset{\sim}{X}$ is $f'(x_i)$.

The estimated response at a point $x\in\chi$ along with its variance is given by

$$\hat{y}(x)=\underline{f}'(x)\ \hat{\beta}$$

$$V(\hat{y}(x))=\sigma^2 f'(x)\ (\underset{\sim}{X}'\ \underset{\sim}{X})^{-1} \underline{f}(x)$$

where $\hat{\beta}$ is least square estimate of β.

Smith [5] proposed the criterion

$$\min_{\{x_i,\ i=1,\ 2,\ \ldots N\}}\ \max_{x\in\chi}\ V(\hat{y}(x))$$

for optimal experimental design when considering the polynomial regression of degree $p-1$ in one variable over the region $\chi=[-1, 1]$. This criterion was later called as G-optimality by Kiefer and Wolfowitz [3]. They extended this criterion in general case to a design measure satisfying

$$\xi(x) \geqslant 0,\ x\in\chi,\ \int_x \xi(dx)=1$$

Thus the variance function of estimated response at a point x is given by

$$d\ (x, \xi)=N\underline{f}'(x)\ (\underset{\sim}{X}'\ \underset{\sim}{X})^{-1} \underline{f}(x).$$

A design with measure ξ^* is said to be G-optimal if

$$\min_{\xi}\ \max_{x\in\chi} d(x, \xi)=\max_{x\in\chi} d(x, \xi^*).$$

A sufficient condition for satisfying above is

$$\max_{x\in\chi} d(x, \xi^*)=p.$$

The G-efficiency of a design ξ is given in comparison to an G-optimal design by $100\ p/\max_{x\in\chi} d(x, \xi)$.

2. The Algoritm

Suppose it is desired to fit the second order response surface given by

$$y=\beta_0+\sum_{i=1}^{v} \beta_i x_i+\sum_{i=1}^{v} \beta_{ii} x_i^2+\sum\sum_{i<j} \beta_{ij} x_i x_j$$

under the experimental region $\chi=\{\underset{\sim}{x} : \underset{\sim}{x}' \underset{\sim}{x} \leqslant 1\}$.

In this case

$$p=(v+1)(v+2)/2.$$

We shall consider only the second order rotatable designs (SORD) introduced by Box and Hunter [1].

The expression of the variance of estimated response at point

$$x_u=(x_{1u}, \ldots\ldots, x_{vu})$$

for the SORD ξ is

$$V(\hat{y}_u)=d(x; \xi)=\theta_0+\theta_1 \underset{\sim}{x}_u' \underset{\sim}{x}_u+\theta_2\left(\underset{\sim}{x}_u' \underset{\sim}{x}_u\right)^2$$

where $\theta_0=(v+2)\lambda_4 \sigma^2/N\Delta$

$$\theta_1=\{(1/\lambda_2)-(2\lambda_2/\Delta)\}\sigma^2/N$$

$$\theta_2=\{1+\left(\lambda_2^2-\lambda_4\right)/\Delta\}\sigma^2/N$$

and

$$\Delta=(v+2)\lambda_4-v\lambda_2^2$$

$$N\lambda_2=\sum_u x_{iu}^2 \qquad \text{for all } i=1, 2, \ldots\ldots, v$$

$$N\lambda_4=\sum_u x_{iu}^2 x_{ju}^2 \qquad \text{for all } i<j=1, 2, \ldots\ldots, v$$

$$\xi(x_u)=1/N, \qquad \text{for all } u=1, 2, \ldots\ldots, N.$$

In order to obtain the factor combination at which the variance of estimated response is maximum, it is easily noticed that the variance expression is a function of $\underset{\sim}{x}' \underset{\sim}{x}$, the square of the distance of point $\underset{\sim}{x}$ from origin. Thus

$$\max V(\hat{y})=\begin{cases} \theta_0 & \text{if } \theta_1+\theta_2 \leqslant 0 \\ \theta_0+\theta_1+\theta_2 & \text{if } \theta_1+\theta_2>0 \end{cases}$$

that is according as the factor levels giving the maximum variance, are at the centre or at the surface of the hypersphere $\underset{\sim}{x}' \underset{\sim}{x}=1$.

The condition $\theta_1+\theta_2 \lesseqgtr 0$ can be expressed as

$$\lambda_4 \lesseqgtr \lambda_4^0$$

where
$$\lambda_4^0 = \left\{ \lambda_2^2 - (\nu+1)\,\lambda_2/2\,(\nu+2) \right\} /2 + \left[\left\{ \lambda_2^2 - (\nu+1)\,\lambda_2/2\,(\nu+2) \right\}^2 + (\nu-1)\,\lambda_2^3/2(\nu+2) \right]^{1/2}$$

In the following, we shall obtain the condition for the choice of the best design. Suppose we have designs D_1 and D_2 in N_1 and N_2 points respectively. There are three different situations for comparisons.

(*i*) Both the λ_4 parameters are less than their respective λ_4^0 parameters. In this case, the design D_1 is preferred to D_2 if

$$\lambda_{42}/\lambda_{41} < \lambda_{22}^2 \Big/ \lambda_{21}^2$$

where λ_{4i} and λ_{2i} are λ_4 and λ_2 parameters of i th design D_i, $i=1, 2$.

In a class of such designs $\{D_j\}$ we shall choose D_i if

$$\min \left\{ \lambda_{2j}^2 / \lambda_{4j} \right\} = \lambda_{2i}^2 / \lambda_{4i}$$

If $\lambda_{2j}=1$ for all j, the design which has the largest λ_4 is preferred.

(*ii*) If λ_4 parameter of D_1 is less than λ_4^0 for the design D_1 and λ_4 of D_2 is greater than λ_4^0 of D_2, then D_1 is preferred to D_2 if

$$N_1 \theta_{01} < N_2\,(\theta_{02}+\theta_{12}+\theta_{22})$$

where $\theta(.)_i$'s are the $\theta(.)$ parameter of D_i, $i=1, 2$.

(*iii*) If λ_4 parameters are greater than λ_4^0 values for both the designs then D_1 is preferred to D_2 if

$$N_1\,(\theta_{01}+\theta_{11}+\theta_{21}) < N_2\,(\theta_{02}+\theta_{12}+\theta_{22})$$

Therefore in order to choose a best design among the set of available designs, one may check one of the above possibilities rather than making ultimate computation of *G*-efficiencies of all the designs as suggested in Lucas [4].

ACKNOWLEDGMENT

The authors are thankful to the referee for his suggestions on an earlier draft.

Summary

An algorithm for the choice of most efficient (according to G-efficiency criterion) design among the set of a several competing designs has been worked out in case of second order rotatable designs.

References

[1] Box, G.E.P. and Hunter, J.S. (1957) : Multifactor experimental designs for exploring response surfaces. Ann. Math. Statist., 28, 195-241.

[2] Kiefer, J. (1959) : Optimal experimental designs. J. Roy Statist. Soc., 20, 272-319.

[3] Kiefer, J. and Wolfowitz. J. (1959) : Optimal designs in regression problems. Ann. Math. Statist. 30, 271-94.

[4] Lucas, J.M. (1976) : A performance comparisons of several types of quadratic response surface designs in symmetric regions. Technometrics, 18, 411-17.

[5] Smith, K. (1918) : On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution observations. Biometrika, 12, 1-85.