Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

THE ROBUST DESIGN OF EXPERIMENTS: A REVIEW

AGNES M. HERZBERG

A review of several optimality criteria for the design of experiments involving quantitative factors for the estimation of unknown parameters in regression problems is given. These include rotatability, *D*-optimality and *G*-optimality criteria. Designs are no longer optimal under these criteria when conditions are suboptimal.

The two main lines of research in the robust design of experiments will be discussed, i. e. (i) the comparison of the performance of an optimal design with respect to a particular criterion under other criteria; (ii) the construction of designs which guard against particular shortcomings.

1. Introduction. The term *robust* was introduced into statistics by Box [5] with regard to statistical tests being valid when some of the assumptions did not hold. Since 1953, there has been a large proliferation of papers on the subject of robustness. Andrews, Bickel, Hampel, Huber, Rogers and Tukey [1] state:

> 'Estimation is the art of inferring information about some unknown quantity on the basis of available data. Typically an estimator of some sort is used. The estimator is chosen to perform well under conditions that are assumed to underly the data. Since these conditions are never known exactly, the estimators must be chosen which are robust, which perform well under a variety of underlying conditions.
>
> The theory of robust estimation is based on the specified properties of specified estimators under specified conditions.'

One wants to determine the interactions of these.

There are broadly two senses in which *robust* is used in the design of experiments. The first is that after one has obtained an optimal design relative to a particular criterion one examines how well the design performs in other respects; see, for example [27; 16; 17; 28]. If there is still some freedom in the construction of the design, one can incorporate further subsidiary optimality requirements; see [22]. The other sense is the construction of designs which guard against particular shortcomings; see for example [8; 20]. These robust designs will be slightly less than optimal under ideal conditions but will be more efficient under more realistic conditions.

Hedayat and John [17] and John [23] have discussed robustness for balanced incomplete block designs. Although very important, this work will not be discussed here.

2. General Comments. The design of experiments involving quantitative factors has been extensively investigated, two main lines of work being that on response surfaces and that on general studies of optimality in designs for estimating regression coefficients; see for example [11; 10; 6] for the first and [15; 12; 24; 25; 26; 29] for the second.

It is well known that caution is necessary in formulating these problems. There are three aspects of primary importance:

(i) the region R of experimentation must be specified. In many cases, the optimal design points will be on the boundary of R;

(ii) the type of model to be fitted has a major effect on the optimal design A choice of the most complex type of model to be fitted has, therefore, to be made. If this is not feasible, the most sensible design is to take all the design points distinct and covering the region R; there is some loss of efficiency if in fact a simple model is adequate;

(iii) the special objectives of the experiment have to be formulated.

Two quite distinct considerations may determine R. Physical constraints may make it impossible or undesirable to work outside certain ranges; or it may be thought that the behaviour of the observations in certain regions of the factors space is of little interest or at least unrepresentative of the behaviour in the main region of concern. In the former case a fairly clearly defined region, often rectangular, is available, but in the second case definition of R is more difficult. Sometimes, however, one tries to measure the factors in units such that a spherical R is reasonably suitable. This suggests three main cases according as R is cuboidal, spherical, and cylindrical. The first two cases have been extensively studied.

Possible objectives of the experiment are numerous. It may be difficult to be very specific and several of the objectives may be involved simultaneously. The main ones are:

(a) the estimation of one or more parameters;

(b) the testing of the adequacy of one or more proposed models, as a guide to the selection of a model;

(c) the discrimination among alternative models;

(d) the estimation of the detailed form of a complex curve or function;

(e) the estimation of the position of and the response at a maximum (or minimum);

(f) the estimation of the line of steepest ascent usually from the center of the design;

(g) the estimation of the properties of the canonical form of the second degree equation;

(h) the estimation of the expected response at one or more points. If the points are inside R, this is a problem of interpolation, whereas if the points are outside R, it is one of extrapolation;

(i) the estimation of the differences of the responses between given points. When the points are close together, there is a connection with (f).

In addition to the previous three central aspects, there are a number of further aspects of importance, including

(iv) the possible presence of additional qualitative factors;

(v) the sequential or nonsequential character of the experiment;

(vi) the error structure, for example blocking, two-way control, splitplot arrangement, etc.;

(vii) the presence of additional constraints, such as the need to have a small number of levels of certain factors.

3. *D*-optimal and *G*-optimal Designs. Let

$$E\{y(\mathbf{x})\} = \mathbf{f}'(\mathbf{x})\boldsymbol{\theta}, \tag{1}$$

where $y(\mathbf{x})$ is an observation at the point $\mathbf{x}$ in the design region, $\mathbf{f}'(\mathbf{x})=\{f_1(\mathbf{x}),\ldots,f_p(\mathbf{x})\}$ and $\mathbf{\theta}'=(\theta_1,\ldots,\theta_p)$, the vector of unknown parameters to be determined by least squares; var $\{y(\mathbf{x})\}=\sigma^2$ and cov $\{y(\mathbf{x}_i),\ y(\mathbf{x}_j)\}=0$ $(i\neq j)$.

Three types of design are of interest:

1) discrete designs; n points $\mathbf{x}_1,\ldots,\mathbf{x}_n$ giving

$$\mathbf{X}=\{\mathbf{f}(\mathbf{x})_n\} \text{ and } \mathbf{X}'\mathbf{X}=\{\sum_{u=1}^{n} f_i(\mathbf{x}_u)f_j(\mathbf{x}_u)\};$$

2) design measures; ξ giving

$$\mathbf{M}(\xi)=\{m_{ij}\},$$

where $m_{ij}=\int_{x\in\chi} f_i(\mathbf{x})f_j(\mathbf{x})\xi(d\mathbf{x})$, χ being the space of interest;

3) associated design measures ξ_n where each point $\mathbf{x}_1,\ldots,\mathbf{x}_n$ has measure or weight $1/n$ and $\mathbf{X}'\mathbf{X}=n\mathbf{M}(\theta_n)$.

The well-known equivalence theorem of Kiefer and Wolfowitz [30] states the following:

Theorem. *The following assertions*

(i) *the design measure* ξ^* *maximizes* $|\mathbf{M}(\xi)|$.

(ii) *the design measure* ξ^* *minimizes* $\max_{\mathbf{x}} d(\mathbf{x},\theta)=\mathbf{f}'(\mathbf{x})\mathbf{M}^{-1}(\xi)f(\mathbf{x})$.

(iii) $\max_{\mathbf{x}} d(\mathbf{x},\xi^*)=p$

are equivalent.

The first assertion defines *D-optimal* designs, $|\mathbf{M}(\xi)|^{-1}$ being proportional to the generalized variances; the second assertation defines *G-optimal* designs, the $d(.,.)$ being related to the variance of the estimated response. For design measures, the two criteria are equivalent, designs satisfying (ii) being easier to determine because of (iii).

Consider the following example. Let

$$\mathsf{E}\{y(\mathbf{x})\}=\theta_0+\theta_1x_1+\theta_2x_2+\theta_{11}x_1^2+\theta_{22}x_2^2+\theta_{12}x_1x_2,$$

and let the design region be the square with vertices $(\pm1,\ \pm1)$. Consider the following three designs. Design I has measure of 0.1458+ at the vertices, 0.08015+ at the midpoints of the edges, and 0.0962+ at the origin. Design II consists of 9 points, one at each vertex, midpoint of the edges and at the origin. Design III consists of 13 points, 2 at each vertex, one at the midpoint of each edge and one at the origin.

For Design I, $|\mathbf{M}(\xi)|=0.011427$, $\max_{\mathbf{x}} d(\mathbf{x},\xi)=6$; for Design II, $|\mathbf{M}(\xi)|=0.0098$, $\max_{\mathbf{x}} d(\mathbf{x},\xi)=7.25$; for Design III, $|\mathbf{M}(\xi)|=0.0113$, $\max_{\mathbf{x}} d(\mathbf{x},\xi)=6.88$. Design I is the *D*-optimal and *G*-optimal design measure, but since the measure is irrational the design cannot be perfomed. Design I serves as a guide. It can be seen that Design III is better than Design II and attains values closest to those of the optimal design.

4. Rotatable Designs. From (1), the least squares estimator of $\mathbf{\theta}$ is

$$\hat{\mathbf{\theta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \tag{2}$$

where $\mathbf{X}=\{\mathbf{f}(\mathbf{x}_u)\}$ and $\mathbf{Y}$ is the $n\times1$ vector of observations. Then, if $\widehat{y}(\mathbf{x})=\mathbf{f}'(\mathbf{x})\hat{\mathbf{\theta}}$,

$$\text{var}\{\widehat{y}(\mathbf{x})\}=\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})\sigma^2, \tag{3}$$

σ^2 being the experimental error variance. Equation (3) will be a function of the distance of the point from the center of the design, i. e

$$\text{var}\{\widehat{y}(\mathbf{x})\} = w(\|\mathbf{x}\|) \tag{4}$$

if and only if

$$(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{N}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{N}, \tag{5}$$

where $\mathbf{N}$ is a particular type of orthogonal matrix; see [10]. A design which satisfies (4) or (5) is called a *rotatable* design. Thus the variance contours of a rotatable design are spherical and the variance of certain $\widehat{\theta}_i$'s are constant no matter what orientation of the design. If the experimenter is not certain in which part of the design region he will be interested, spherical variance contours appear to be advisable; see, for example [10]. These designs could be considered as being variance robust.

5. Model Robustness. Box and Draper [6; 7] considered designs that minimize the mean square error over the spherical region of the factor space. They minimized, over the spherical region R, the function

$$J = \frac{n\Omega}{\sigma^2}\int_R \mathsf{E}\{\widehat{y}(\mathbf{x}) - \eta(\mathbf{x})\}^2 d\mathbf{x} = \frac{n\Omega}{\sigma^2}\int_R \text{var}\{\widehat{y}(\mathbf{x})d\mathbf{x} + \frac{n\Omega}{\sigma^2}\int_R [\mathsf{E}\{y(\mathbf{x})\} - \eta(\mathbf{x})]^2 d\mathbf{x},$$

where $\widehat{y}(\mathbf{x})$ is the least squares estimate of the assumed polynomial model, $\eta(\mathbf{x})$ is the true model, $\Omega^{-1} = \int_R d\mathbf{x}$, n is the number of design points and σ^2 is the experimental error variance. The size of J arises partly from bias due to the use of a polynomial of too small degree and partly from random errors. When bias alone is considered, the bias is minimized by rotatable designs. Draper and Lawrence [13; 14] examined the situation when the region is cuboidal.

Kiefer and Studden [28] compared designs for situations where the model is of the form

$$\mathsf{E}\{y(\mathbf{x})\} = \sum_{i=0}^{n} \theta_i x^i,$$

the design region is $[-1, 1]$ and the optimal design measure, ξ_n, has n points of support. They discussed optimal n point design measures for interpolation and extrapolation and compared these to the limiting case when $n \to \infty$. It is shown that in some situations it may be better when the model is of order n to use ξ_m, where $m > n$.

6. The Variance of the Difference between Two Responses. The experimenter may not always be interested in the estimated response at one point but changes of response, for example, in the difference between the estimated responses at two points. Herzberg [19] considered the behaviour of the variance function of the difference between two estimated responses. It was shown that for rotatable designs, the variance function of the difference between two estimated responses is a function only of the distances of the points from the origin of the design and the angle subtending them at the origin. Box and Draper [9] have extended and unified these results.

7. Measures of Design Robustness. Herzberg and Andrews [20] introduced robust measures for situations, where outliers, missing values or non-Gaussian distributions are contemplated.

Let $\alpha(\mathbf{x})$ be the probability of losing an observation at the point $\mathbf{x}$, the losses at different points being independent. Let $\mathbf{D} = \{d_{ii}\}$, a diagonal matrix, where

$$d_{ii}=\begin{cases}0 \text{ with probability } \alpha(\mathbf{x}_i),\\ 1 \text{ with probability } 1-\alpha(\mathbf{x}_i),\end{cases}$$

the value 0 being associated with a missing observation. This, of course, does not explain missing observations, but is only a formulation to investigate their potential influence on experimental design.

If a sufficient number of observations are missing, not all of the elements of $\boldsymbol{\theta}$ given in (1) may be estimated. This occurs when $|\mathbf{X}'\mathbf{D}\mathbf{X}|=0$. The probability that this occurs is

(6) $$\mathrm{pr}(|\mathbf{X}'\mathbf{D}\mathbf{X}|=0),$$

which may be used to compare designs. This measure is called the *probability of breakdown* of the design. Designs robust under this measure have a small value for (6).

Another measure of robustness is

(7) $$\mathrm{E}(|\mathbf{X}'\mathbf{D}\mathbf{X}|^{1/p}),$$

which is related to the D-optimality criterion. Designs robust under this measure have (7) large.

Another measure of robustness and some examples are given in [20]. Further examples and computational methods are given in [2]. Herzberg and Andrews [21] compare the robustness of chain block and coat-of-mail designs.

Box and Draper [8] introduced a criterion for the construction of designs to minimize the effect of outliers in the least squares estimate of the response function. For the comparison of designs, the value of the population variance of the variances of the estimated response at the design points is used implying the smaller the value the more robust the design.

8. Robustness against Autocorrelation in Time. The assumption of $\mathrm{cov}\{y(\mathbf{x}_i), y(\mathbf{x}_j)\}=0$ $(i\neq j)$ is very often artificial. Bickel and Herzberg [3] assume that $\mathrm{cov}\{y(\mathbf{x}_i), y(\mathbf{x}_j)\}=v_{ij}\sigma^2$ and the variance-covariance matrix of the observations is $\mathbf{V}\sigma^2$, where $\mathbf{V}=\{v_{ij}\}$. They find one-dimensional optimal designs for particular situations when the number of observations n tends to infinity and $\mathbf{V}$ is unknown but of an assumed form. Examples and further results are given in [4].

9. Conclusion. A review of several optimality criteria for the design of experiments has been given. Some of these designs possess certain robust qualities. Measures of robustness of designs are introduced in order that designs may be compared. It is very important also in this connection to examine the results of Kiefer [27] and Galil and Kiefer [16; 17], where optimal designs relative to a particular procedure are examined for their performance with respect to other criteria.

Acknowledgement. I am grateful to Professor D. R. Cox for permission to include some joint work here. This paper was written while the author was visiting the Mathematics Research Center at the University of Wisconsin. The work was supported under contract No. DAAG 29-80-c-0041.

REFERENCES

1. D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, J. W. Tukey. Robust Estimates of Location: Survey and Advances. Princeton, 1972.
2. D. F. Andrews, A. M. Herzberg. The robustness and optimality of response surface designs. *J. Statist. Planning and Inference*, **3**, 1979, 249—257.
3. P. J. Bickel, A. M. Herzberg. Robustness of design against autocorrelation in time I: Asyptotic theory, optimality for location and linear regression. *Ann. Statist.*, **7**, 1979, 77—95.
4. P. J. Bickel, A. M. Herzberg, M. Schilling. Robustness of design against autocorrelation in time II: Numerical results for the first order autoregressive process. *J. Amer. Statist. Assoc.*, (to appear).
5. G. E. P. Box. Non-normality and tests on variances. *Biometrica*, **40**, 1953, 318—335.
6. G. E. P. Box, N. R. Draper. A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.*, **54**, 1959, 622—654.
7. G. E. P. Box, N. R. Draper. The choice of a second order rotatable design. *Biometrica* **50**, 1963, 335—352.
8. G. E. P. Box, N. R. Draper. Robust design. *Biometrica*, **62**, 1975, 347—352.
9. G. E. P. Box, N. R. Draper. The variance function of the difference between two estimated responses. *J. R. Statist. Soc. B*, **42**, 1980, 79—82.
10. G. E. P. Box, J. S. Hunter. Multifactor experimental designs for exploring response surfaces. *Ann. Math. Statist.*, **28**, 1957, 195—241.
11. G. E. P. Box, K. P. Wilson. On the experimental attainment of optimum conditions, (with discussion). *J. R. Statist. Soc. B*, **13**, 1951, 1—45.
12. H. Chernoff. Locally optimal designs for estimating parameters. *Ann. Math. Statist.*, **25**, 1953, 123—130.
13. N. R. Draper. W. E. Lawrence Designs which minimize model inadequacies: cuboidal regions of interest. *Biometrica*, 1965, **52**, 111—118.
14. N. R. Draper, W. E. Lawrence. The use of second-order 'spherical' and 'cuboidal' designs in the wrong regions. *Biometrica*, **53**, 1966, 596—599.
15. G. Elfving. Optimal allocation in linear regression theory. *Ann. Math. Statist.*, **23**, 1952, 255—262.
16. Z. Galil, J. Kiefer. Comparison of designs for quadratic regression on cubes. *J. Statist. Planning and Inference*, **1**, 1977, 121—132.
17. Z. Galil, J. Kiefer. Comparison of rotatable designs for regression on balls, I (Quadratic). *J. Statist. Planning and Inference*, **1**, 1977, 27—40.
18. A. Hedayat, P. W. M. John. Resistant and susceptible BIB designs. *Ann, Statist.*, **2**, 1974, 148—158.
19. A. M. Herzberg. The behaviour of the variance function of the difference between two estimated responses. *J. R. Statist. Soc. B*, **29**, 1967, 174—179.
20. A. M. Herzberg, D. F. Andrews. Some considerations in the optimal design of experiments in non-optimal situations. *J. R, Statist. Soc. B*, **38**, 1976, 284—289.
21. A. M. Herzberg, D. F. Andrews. The robustness of chain block designs and coat-of-mail designs. *Comm. Statist. A*, **7**, 1978, 479—485.
22. A. M. Herzberg, D. R. Cox. Some optimal designs for interpolation and extrapolation *Biometrica*, **59**, 1972, 551—561.
23. P. W. M. John. Robustness of balanced incomplete block designs. *Ann. Statist.*, **4**, 1976 960—962.
24. J. Kiefer. Optimum experimental designs (with discussions). *J. R. Statist. Soc. B*, **21**, 1959, 272—319.
25. J. Kiefer. Optimum designs in regression problems, II, *Ann. Math. Statist.*,**32**, 1961,293—325.
26. J. Kiefer. Two more criteria equivalent to *L*-optimality of designs. *Ann. Math. Statist.* **33**, 1962, 792—796.
27. J. Kiefer. Optimal design: Variation in structure and performance under change of criterion. *Biometrica*, **62**, 1975, 277—288.
28. J. Kiefer, W. J. Studden. Optimal designs for large degree polynomial regression. *Ann. Statist.*, **4**, 1976, 1113—1123.
29. J Kiefer, J. Wolfowitz. Optimum designs in regression problems. *Ann. Math. Statist.*, **30**, 1959, 271—294.
30. J. Kiefer, J. Wolfowitz. The equivalence of two extremum problems. *Canad. J, Math.*, **12**, 1960, 363—366.

Imperial College of Science and Technology
University of London London, England

Received 7. 7. 1981