Title
Comparing Robust Properties of A, D, E and G-Optimal Designs

Author
Weng-Kee Wong

Publication Date
2011-10-24

Computational Statistics & Data Analysis 18 (1994) 441–448
North-Holland

Comparing robust properties of A, D, E and G-optimal designs

Weng Kee Wong
Department of Biostatistics, University of California at Los Angeles, Los Angeles, CA, USA

Received February 1993
Revised June 1993

Abstract: We examine the A, D, E and G-efficiencies of using the optimal design for the polynomial regression model of degree k when the hypothesized model is of degree j and $1 \leqslant j \leqslant k \leqslant 8$. The robustness properties of each of these optimal designs with respect to the other optimality criteria are also investigated. Relationships among these efficiencies are noted and practical implications of the results are discussed. In particular, our numerical results show E-optimal designs possess several properties not shared by the A, D and G-optimal designs.

Keywords: Continuous designs; A, D, E and G-efficiencies; Homoscedasticity.

1. Introduction

The purpose of this paper is to make a numerical comparison of the efficiencies of different types of optimal designs under various model assumptions. This is an important consideration because most optimal designs are very model dependent and the true model is usually unknown in practice. In addition, we also evaluate the robustness properties of the optimal designs under different optimality criteria. The importance of using a design that is deemed adequate for several optimality criteria cannot be overemphasized since optimality with respect to any one criterion usually but represents an approximation to some vague notion of "goodness", see Williams (1958) and especially Kiefer (1975) for a more thorough discussion on this subject. Many experimental designs and the subsequent analyses do depend on the model assumptions (Box and Hay, 1953) and it is therefore important to assess the adequacy of a design under different model assumptions and different optimality criteria.

All our comparisons are made under the assumption that the true model $f_j(x)$ is a polynomial of degree j, $1 \leqslant j \leqslant 8$, the response variable is univariate

Correspondence to: W.K. Wong, Ph.D., Department of Biostatistics, UCLA School of Public Health, 10833 Le Conte Avenue, Los Angeles, CA 90024-1772, USA.

SSDI 0167-9473(93)E0029-4

and the errors in the observations are uncorrelated with zero mean and constant variance. Without loss of generality, the design space Ω is taken to be $[-1, 1]$. The choice of this setting is the same as in Kussmaul (1969), Kendall and Stuart (1968), among others, who suggested that using the optimal design for $f_k(x)$ may be a good idea for the assumed model $f_j(x)$, $k > j$. Doing so will at least enable the experimenter to perform a lack of fit test.

Only continuous designs are considered here, i.e., we treat designs as probability measures on Ω. If ξ is a design with mass m_i at $x_i \in \Omega$, $i = 1, 2, \ldots, k$, ($\sum_{i=1}^{k} m_i = 1$) and n observations are planned, then approximately nm_i observations will be taken at x_i, $i = 1, 2, \ldots, k$. In what follows, we shall assume that n is large.

Following standard optimal design theory (Fedorov, 1972), the amount of information contained in a continuous design ξ is measured by its information matrix:

$$M_j(\xi) = \int_\Omega f_j(x) f_j^{\mathrm{T}}(x) \xi(\mathrm{d}x). \tag{1.0}$$

Many practical objectives in an experiment can be expressed in terms of the information matrix. For example, if the true underlying model is $f_j(x)$ and the goal of the experiment is to estimate the parameters as accurately as possible, then the optimal continuous design is the D-optimal design ξ_j defined by $|M_j(\xi_j)| = \max_{\xi \in \Xi} |M_j(\xi)|$, where $|A|$ is the determinant of the matrix A and Ξ is the set of all continuous designs on Ω. Another is the G-optimality criterion, which minimizes the maximum variance of the estimated response surface across Ω, i.e., a G-optimal design minimizes the quantity $\max_{x \in \Omega} d_j(x, \xi)$ over Ξ, where $d_j(x, \xi) = f_j^{\mathrm{T}}(x) M_j^{-1}(\xi) f_j(x)$. This design is particularly useful when there is interest in estimating the entire response surface. Under the assumption of homoscedasticity, it is known that ξ_j is also G-optimal (Kiefer–Wolfowitz Theorem, 1960).

Two other optimality criteria are included in our study: the A and E-optimality criteria; for the model $f_j(x)$, the A-optimal design is denoted by $\xi_{\mathrm{A},j}$ and the E-optimal design is denoted by $\xi_{\mathrm{E},j}$. Together, these four comprise perhaps the most interpretable alphabetical optimality criteria (Kiefer, 1975). Further details on the motivation and uses of these criteria can be found in Kiefer (1985).

Virtually, all designs are ranked on the basis of their efficiencies. When a design ξ has a nonsingular information matrix, its G, D, A and E-efficiencies are respectively given by

$$G_j(\xi) = (j+1) / \max_{x \in \Omega} d_j(x, \xi),$$

$$D_j(\xi) = |M_j^{-1}(\xi_j) M_j(\xi)|^{1/(j+1)},$$

$$A_j(\xi) = \operatorname{tr} M_j^{-1}(\xi_{\mathrm{A},j}) / \operatorname{tr} M_j^{-1}(\xi)$$

and

$$E_j(\xi) = \lambda_{\max}\big(M_j^{-1}(\xi_{\mathrm{E},j})\big) / \lambda_{\max}\big(M_j^{-1}(\xi)\big). \tag{1.1}$$

Here tr M and $\lambda_{max}(M)$ denote the trace and maximum eigenvalue of the square matrix M respectively. The interpretation of these efficiencies is they measure the number of times the experiment needs to be replicated in order to have the same criterion value as the optimal design. Clearly, the efficiency of each of the optimal designs relative to its own optimality criterion is equal to one.

G and E-optimality are special cases of mini–max type of criterion and both reduce to the problem of finding a design that minimizes $\max_{t \in T} t^T M_j^{-1}(\xi)t$ for some set T. It is easy to show that if $T = \{t \mid t = f_j(x),\ x \in \Omega\}$, we have G-optimality, and if $T = \{t \in R^{j+1} \mid t^T t = 1\}$, E-optimality obtains. For a general approach of constructing such type of designs, see Wong (1992). It should also be noted that unlike D or G-optimal designs, both A and E-optimal designs are not invariant under nonsingular linear transformation.

The optimality of the designs ξ_j, $\xi_{A,j}$ and $\xi_{E,j}$ can be verified using the checking conditions described in Kiefer (1975) and Kiefer (1985, p. 319). Computation of the D and G-efficiencies are straightforward since formulae for ξ_j are well known (see Fedorov, 1972, p. 89). However, the calculation of the A and E-efficiencies are made possible only from recently established results concerning the theoretical A and E-optimal design for $f_j(x)$ for $j \geqslant 3$. A-optimal designs were found by Pukelsheism and Torsney (1991) while analytic formulae for E-optimal designs were given in Pukelsheim and Studden (1991). The A-efficiencies in Table 2.2 are calculated with the help of Table 3 in Pukelsheim and Torsney (1991), and the E-efficiencies are found by first enumerating the

Table 2.1
E-optimal designs on $[-1, 1]$

Degree of poly-nomial	E-optimal support points (upper row) and their E-optimal weights (lower row)									Optimal value E-criterion
1	−1	1								1
	0.500	0.500								
2	−1	0	1							5
	0.200	0.600	0.200							
3	−1	−0.5	0.5	1						25
	0.127	0.373	0.373	0.127						
4	−1	−0.707	0.000	0.707	1					129
	0.093	0.248	0.318	0.248	0.093					
5	−1	−0.809	0.309	0.309	0.809	1				681
	0.074	0.180	0.246	0.246	0.180	0.074				
6	−1	−0.866	−0.500	0.000	0.500	0.866	1			3653
	0.061	0.141	0.189	0.218	0.189	0.141	0.061			
7	−1	−0.901	−0.624	−0.223	0.223	0.624	0.901	1		19825
	0.052	0.116	0.149	0.183	0.183	0.149	0.116	0.052		
8	−1	−0.924	−0.707	0.383	0.000	0.383	0.707	0.924	1	108545
	0.045	0.098	0.122	0.152	0.167	0.152	0.122	0.098	0.045	

Table 2.2
A, D, E and G-efficiencies $1 \leqslant j \leqslant k \leqslant 8$

Degree of polynomial	Type of efficiency	k $(1 \leqslant j \leqslant k \leqslant 8)$							
		1	2	3	4	5	6	7	8
$j=1$	E	1.000	0.400	0.440	0.434	0.430	0.428	0.426	0.424
	A	1.000	0.667	0.621	0.610	0.604	0.601	0.598	0.597
	G	1.000	0.800	0.750	0.727	0.714	0.706	0.700	0.696
	D	1.000	0.817	0.775	0.756	0.745	0.739	0.734	0.730
$j=2$	E		1.000	0.439	0.503	0.502	0.503	0.504	0.504
	A		1.000	0.640	0.612	0.611	0.611	0.610	0.610
	G		1.000	0.818	0.750	0.714	0.692	0.677	0.667
	D		1.000	0.865	0.828	0.809	0.798	0.790	0.784
$j=3$	E			1.000	0.434	0.504	0.500	0.500	0.500
	A			1.000	0.623	0.603	0.602	0.602	0.602
	G			1.000	0.842	0.769	0.727	0.700	0.681
	D			1.000	0.895	0.863	0.847	0.836	0.829
$j=4$	E				1.000	0.430	0.505	0.500	0.500
	A				1.000	0.614	0.597	0.597	0.597
	G				1.000	0.862	0.790	0.745	0.714
	D				1.000	0.915	0.888	0.873	0.863
$j=5$	E					1.000	0.428	0.506	0.500
	A					1.000	0.609	0.594	0.594
	G					1.000	0.878	0.808	0.762
	D					1.000	0.928	0.904	0.891
$j=6$	E						1.000	0.426	0.507
	A						1.000	0.608	0.593
	G						1.000	0.891	0.824
	D						1.000	0.945	0.917
$j=7$	E							1.000	0.424
	A							1.000	0.601
	G							1.000	0.901
	D							1.000	0.945

optimal designs and then evaluating (1.1). Since E-optimal designs for polynomial regression of degrees higher than 2 appears to be generally unknown, they are tabulated in Table 2.1. They are derived as follows: Let $c^{\mathrm{T}} = (c_0, c_1, \ldots, c_j)$ be the coefficient vector of the Chebyshev polynomial of degree j, $s_i = \cos\{(j-i)\pi/j\}$, $i = 0, 1, 2, \ldots, j$ and $u_0, u_1, \ldots, u_j$ be the solution of the system of linear equations $\Sigma_{i=0}^{j} u_i f_j(s_i) = c$. Then, the E-optimal design $\xi_{\mathrm{E},j}$ for $f_j(x)$ is supported at s_t with mass w_t given by $(-1)^{j-t} u_t / \Sigma_{i=0}^{j} c_i^2$, $t = 0, 1, 2, \ldots, j$, see Pukelsheim and Studden (1991).

In the next section, we demonstrate that a design that is deemed adequate under one criterion can do poorly in terms of another. While this message is not new at all, it appears there is no work in the literature addressing this issue in terms of these commonly used criteria and in the setting considered here. Section 3 concludes with general guidelines on the choice of these optimal designs among competing designs.

2. Results

Table 2.2 lists the A, D, E and G-efficiencies for $1 \leqslant j \leqslant k \leqslant 8$. Notice that, as expected, for fixed j, $D_j(\xi_k)$, $G_j(\xi_k)$ and $A_j(\xi_{A,k})$ are all non-increasing functions of k. Surprisingly, $E_j(\xi_{E,k})$ does not have this monotonic property and in fact, for $1 \leqslant j < k \leqslant 8$, it remains remarkably stable, averaging an value of only 0.47. It follows that if the true model is $f_j(x)$, the E-efficiency of ξ_k is always unacceptably low and insensitive to the value of k, so long as $k > j$. Consequently, if the experimenter is unsure of the underlying model, he should be very cautious about using E-optimality since a slightly misspecified model can result in severe loss in efficiency of the design. On the other hand, the G and D-optimal designs seem to offer fairly good protection against small departures from the true model, averaging a G or D-efficiency of 0.8 or more, but this is not true for A-optimal designs. For instance using $\xi_{A,j+1}$ when the true model is $f_j(x)$ results in an A-efficiency of about 0.615 for $1 \leqslant j \leqslant 8$. (Here small departure is taken to mean $k - j$ is less than 2.)

Another property of E-optimal designs not shared by the A, D and G-optimal designs is seen from Table 2.2. It shows a rather counter-intuitive relationship,

$$E_j(\xi_{E,s}) > E_j(\xi_{E,j+1}) \quad \text{for } 8 \geqslant s > j + 1 \geqslant 2.$$

The implication of this seems to be that is one assumes $f_j(x)$ is the true model, then using $\xi_{E,k}$, $k > j + 1$ would provide a higher E-efficiency than using $\xi_{E,j+1}$. For example, if the assumed model is $f_1(x)$ then using the design $\xi_{E,k}$ $(k > 2)$ will result in higher E-efficiency than $\xi_{E,2}$ which is "closer" than $\xi_{E,1}$. There is however, a consistent pattern from the table, namely for $1 \leqslant j \leqslant k \leqslant 8$,

$$D_j(\xi_k) \geqslant G_j(\xi_k) \geqslant A_j(\xi_{A,k}) \geqslant E_j(\xi_{E,k}).$$

The ordering in this string of inequalities is quite pleasing for at least two reasons: (1) the D, A and E-optimality criteria all belong to Kiefer's Φ_p-optimality criterion (1975) and they correspond to $p = 0$, -1 and $-\infty$ respectively in that order, and (2) it is in accordance with the mathematical difficulty involved in studying these designs. The G-optimality does not belong to the class of the Φ_p-optimality criterion but the above relationship suggests it may correspond to some p between 0 and -1. Note that the above relationship between the D and G-efficiencies is a special case of Atwood's inequality (1969).

Kiefer (1975) advocated that any recommended designs should be compared on the basis of several criteria of goodness. Table 2.3 reports the changes in efficiencies of the optimal designs when the optimality criterion is varied. Again, all the models are polynomial regression models of degrees ranging from 1 to 8.

3. Discussion

The D-optimal designs are perhaps the most frequently used designs but our computation shows they do not fair well in terms of E-efficiencies; they average

Table 2.3
Efficiencies of optimal designs under change of optimality criteria

	j	A-eff.	D-eff.	E-eff.	G-eff.
$\xi_{A,j}$	2	1	0.946	0.955	0.750
	3	1	0.916	0.968	0.600
	4	1	0.907	0.968	0.520
	5	1	0.906	0.970	0.480
	6	1	0.905	0.970	0.455
	7	1	0.911	0.968	0.440
	8	1	0.912	0.971	0.423
ξ_j	2	0.889	1	0.731	1
	3	0.853	1	0.745	1
	4	0.839	1	0.739	1
	5	0.831	1	0.735	1
	6	0.829	1	0.733	1
	7	0.824	1	0.729	1
	8	0.817	1	0.708	1
$\xi_{E,j}$	2	0.960	0.864	1	0.600
	3	0.969	0.863	1	0.508
	4	0.972	0.868	1	0.465
	5	0.970	0.876	1	0.442
	6	0.970	0.880	1	0.427
	7	0.975	0.890	1	0.416
	8	0.971	0.897	1	0.405

about 0.73. However, they have moderately high A-efficiencies (about 0.83), which decrease as the degree of the polynomial regression increases. The G-efficiencies of D-optimal designs are all equal to one by the Kiefer–Wolfowitz Theorem (1960). In contrast, the A and D-efficiencies of E-optimal designs seem quite insensitive to changes in the model specification. They perform exceptionally well, averaging about 0.97 for A-efficiencies and 0.87 for D-efficiencies. The G-efficiencies of E-optimal designs tend to be low averaging less than 0.5 and this is true for A-optimal designs too. The highest efficiency is only 0.75 when the A-optimal design is used for the G-optimality criterion in the quadratic model. Thus, G-optimality should be used sparingly since most of the other optimal designs do not fair well in terms of G-efficiencies at all. A-optimal designs have uniformly high E-efficiencies and they outperform the E-optimal designs in terms of D-efficiencies.

A formal mathematical justification of the findings here appears difficult, partly because A and E-optimal designs cannot be described in a nice closed form like those of D and G-optimal designs. However, a general result relating the D, A and E-efficiencies can be derived from the arithmetic-geometric-harmonic mean inequality: if the regression function has p components, it is straightforward to show that if

$$1/\lambda_{\max}(M(\xi_D)) \geqslant \frac{1}{p}\operatorname{tr} M^{-1}(\xi_A) \geqslant \lambda_{\max}(M^{-1}(\xi_E)), \tag{3.0}$$

i.e., if the average of the eigenvalues of an A-optimal design is not too extreme, then

$$D(\xi) \geqslant A(\xi) \geqslant E(\xi) \tag{3.1}$$

for any linear models and any design ξ whose information matrix is nonsingular. For simple linear regression model with $j = 1$, (3.0) is trivially true and thus (3.1) holds. More generally, one can show if any part of (3.0) is satisfied, then the corresponding part of (3.1) is true. Unfortunately, condition (3.0) appears to be too stringent for most applications; for $j = 2$, the numbers in (3.0) are 0.657671, 2.6667 and 5 from left to right so that the direction of the inequality in (3.0) is completely reversed. On the other hand, this is perhaps not unexpected since validity of (3.1) would imply that a D-optimal design is also A and E-optimal as well. Likewise, an A-optimal design must also be E-optimal.

If one considers a design as (very) efficient if it has an efficiency of (0.9) 0.8 or higher, we may conclude from this study that (i) D-optimal designs are efficient in terms of A-efficiencies, very efficient in terms of G-efficiencies but they are not efficient in terms of E-efficiencies, (ii) E-optimal designs are very efficient in terms of A-efficiencies, efficient in terms of D-efficiencies but they are not efficient in terms of G-efficiencies and, (iii) A-optimal designs are very efficient in terms of both D and E-efficiencies but they are not efficient in terms of G-efficiencies. The G-optimality criterion seems to be compatible with D-optimality only.

Acknowledgement

Thanks go to Wen-Yin Lin for help with the computation.

References

Atwood, C.L., Optimal and efficient designs of experiments, *Ann. Math. Statist.*, **40**(2) (1969) 1570–1602.

Box, G.E.P. and W.A.A. Hay, Statistical design for the efficient removal of trends occurring in a comparative experiment with an application in biological assay, *Biometrics*, **9**(3) (1953) 304–319.

Fedorov, V.V., *Theory of optimal experiments*, translated and edited by W.J. Studden and E.M. Klimko, (Academic Press, New York, 1972).

Kendall, M.G. and A. Stuart, *The advanced theory of statistics*, Vol. 3 (Hafner Publishing Co., New York, 1968).

Kiefer, J. and Wolfowitz., The equivalence of two extremum problems, *Canad. J. Math.*, **12** (1960) 363–366.

Kiefer, J., Optimal design: Variation in structure and performance under change of criterion, *Biometrika*, **62**(2) (1975) 277–288.

Kiefer, J., *Jack Carl Kiefer collected papers III: Design of experiments* (Springer-Verlag, New York, 1985).

Kussmaul, K., Protection against assuming the wrong degree in polynomial regression, *Technometrics*, **11**(4) (1969) 677–682.

Pukelsheim, F and B. Torsney, Optimal weights for experimental designs on linearly independent support points, *Ann Statist.* **19**(3) (1991) 1614–1625.

Pukelsheim F and W.J. Studden, E-optimal designs for polynomial regression, Mimeo #91-24 (Dept. of Statistics, Purdue University, West Lafayette, IN, 1991).

Williams, E.J., Optimum allocation for estimation of polynomial regression, *Biometrics*, **14**(4) (1958) 573–574.

Wong, W.K., A unified approach to the construction of minimax designs, *Biometrika*, **79**(3) (1992) 611–620.