OPTIMAL DESIGNS FOR QUADRATIC REGRESSION

Franz PREITSCHOPF and Friedrich PUKELSHEIM

Institut für Mathematik der Universität Augsburg, D-8900 Augsburg, Federal Republic of Germany

1. Introduction

The abundance of optimality criteria which are available in experimental design theory calls for a qualitative comparison of how various criteria compare, for various parameter sets. Galil and Kiefer (1977), Pukelsheim (1980, Section 6), and Studden (1980) are some of the papers which present results along these lines. That work is complemented here by a somewhat detailed report of what happens in quadratic regression on the interval $[-1, +1]$. This model lends itself to a particularly simple study because the set of all designs may be reduced through symmetry to a complete class which is one-dimensional, depending on a single weight $\alpha \in [0, 1]$ only.

The optimality criteria to be mainly considered are the p-means, with $-\infty \leq p \leq 1$, of the eigenvalues of the information matrices for the parameters of interest. When p varies between $-\infty$ and 0 this p-mean optimality coincides with Kiefer's (1975, p. 279) Φ_p-optimality, and covers the classical criteria such as D-, A-, and E-optimality. The case $p = 1$ corresponds to trace-optimality which is instrumental in Kiefer's (1975) concept of universal optimality. Indeed, the domain of variation of

p extends to also include the full interval from 0 to 1, cf. Pukelsheim (1980, p. 342). It is there that in quadratic regression the optimal weight function $\alpha(p)$ and the optimal information value $v(p)$ grow fastest, and have points of inflection.

2. Quadratic regression model

Let the observation Y depend on the design variable x through

$$Y(x)=\beta_0+\beta_1 x+\beta_2 x^2+\sigma e, \quad x\in[-1;1],$$

where e is the random error with mean 0 and variance 1 and $\sigma>0$ is the unknown scaling factor, while the quadratic expression forms the mean regression. Repeated observations are assumed to be independent, as usual. By abuse of language we call $\beta_0, \beta_1, \beta_2$ the *constant*, *linear*, *quadratic parameter*, respectively. Interest is in finding optimal designs for any given subset of these parameters.

A design ξ is a discrete probability distribution on the experimental region $[-1, +1]$ and determines allocation and proportion of the observations. A particular role is played by the *symmetric three-point designs*

$$\xi_\alpha(-1)=\xi_\alpha(+1)=\tfrac{1}{2}\alpha, \quad \xi_\alpha(0)=1-\alpha, \qquad \alpha\in[0, 1].$$

It is not hard to show that for every design ξ we can find a design ξ_α which is at least as good as ξ, see Kiefer (1959, Lemma 3.5) or Giovagnoli, Pukelsheim and Wynn (1987, Section 4). Therefore it suffices to determine the optimal member within the one-dimensional class of symmetric three-point designs, parametrized by α. This reduction greatly facilitates the problem.

3. Optimality results for means of information eigenvalues

Let the parameter system of interest be represented in the form

$$K'\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix},$$

with an appropriate $3\times s$ matrix K. The p-mean criterion, with $-\infty\leq p\leq 1$, maximizes information for the parameters of interest through the function $j_p(J(M_\alpha))$, where $j_p(J)$ is the generalized mean of order p of the eigenvalues of J, and $J(M)$ is the information matrix for the parameters of interest (i.e. $J(M)=(K'M^-K)^{-1}$ in case of identifiability and $J(M)=0$ otherwise), and M_α is the moment matrix of the design ξ_α as given in Table 2. The *optimal weight* will be denoted by $\alpha(p)$, and the value

$$v(p)=j_p(J(M_{\alpha(p)}))=\sup_{\alpha\in[0;1]} j_p(J(M_\alpha))$$

then gives the *optimal information* for the parameters of interest. Our results pertain to the behaviour of the optimal weight $\alpha(p)$ and the optimal information $\upsilon(p)$, as functions of the criterion parameter p.

Only for three subsets of the parameters will the optimal designs depend on the criterion parameter p: the constant-linear-quadratic, the linear-quadratic, and the constant-quadratic parameters.

Figure 1 shows that the optimal weight functions α look almost constant between $-\infty$ and -6, increase moderately towards 0, and show a steeper ascent in the interval between 0 and 1. For the linear-quadratic parameters the optimal weight function α is seen to be convex. For the constant-linear-quadratic parameters the function α has a point of inflection at $\alpha(0.450)=0.872$, and for the constant-quadratic parameters a point of inflection is $\alpha(0.608)=0.822$. The constant-linear-quadratic case ends in an almost constant piece for $0.8 \leq p \leq 1$. For $p=1$ no optimal

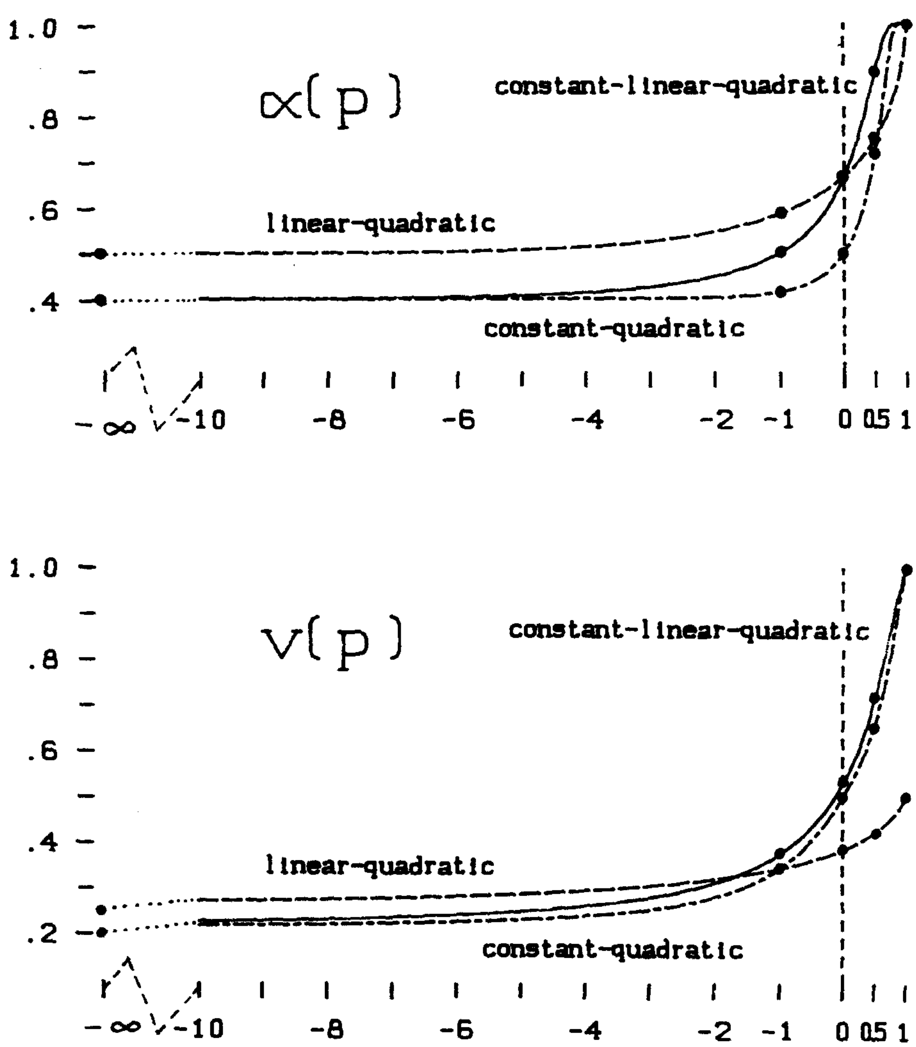

Fig. 1. Graphs of optimal weight α and optimal information υ for three parameter subsets in quadratic regression, as functions of the criterion parameter $p \in [-\infty, +1]$ when the objective is to maximize the p-mean of the information eigenvalues. Dots indicate exact results. For $p=1$ no optimal weights exist whence $\alpha(1)$ are limiting values.

design exists for either of the three cases, due to lack of identifiability. The D-criterion is particularly pleasing since the weight $\alpha = \frac{2}{3}$ is optimal for both the constant-linear-quadratic, and the linear-quadratic parameters.

The optimal information values υ in Figure 1 show a similar behaviour. However, when p tends to $-\infty$ convergence is much slower. At the other end of the scale the three υ-curves fail to end in a common value for $p = 1$. Again there are points of inflection, in the constant-linear-quadratic case the coordinates are $\upsilon(0.710) = 0.836$, in the constant-quadratic case $\upsilon(0.828) = 0.867$.

Table 1 presents numerical values of $\alpha(p)$ and $\upsilon(p)$ for selected values of p. Exact values were obtained for $p = -\infty$, -1, 0, 0.5, and 1, as well as for the remaining subsets of the constant-linear parameters and the three single parameters.

We briefly comment on the derivation of these results. With fixed parameter $p \in [-\infty, +1]$ the problem reads:

Table 1

Numerical and exact values of optimal weight α and optimal information υ for all parameter subsets in quadratic regression. Only for the first three parameter subsets these values depend on the criterion parameter $p \in [-\infty, -1]$ when the objective is to maximize the p-mean of the information eigenvalues. For $p = 1$ braces indicate limiting cases where no optimal design exists

mean	constant-linear-quadratic		linear-quadratic		constant-quadratic	
p	$\alpha(p)$	$\upsilon(p)$	$\alpha(p)$	$\upsilon(p)$	$\alpha(p)$	$\upsilon(p)$
$-\infty$	0.400 = 2/5	0.200 = 1/5	0.500 = 1/2	0.250 = 1/4	0.400 = 2/5	0.200 = 1/5
−10	0.400	0.223	0.500	0.268	0.400	0.214
−5	0.407	0.248	0.507	0.285	0.400	0.230
−3	0.425	0.278	0.525	0.304	0.400	0.252
−2.5	0.435	0.292	0.534	0.311	0.401	0.263
−2	0.449	0.310	0.547	0.319	0.402	0.279
−1.5	0.468	0.336	0.563	0.330	0.406	0.304
−1	0.500 = 1/2	0.375 = 3/8	$0.586 = 2-\sqrt{2}$	$0.343 = 6-4\sqrt{2}$	$0.414 = \sqrt{2}-1$	$0.343 = 6-4\sqrt{2}$
−0.5	0.555	0.434	0.618	0.361	0.436	0.405
0	0.667 = 2/3	$0.529 = 4^{1/3}/3$	0.667 = 2/3	$0.385 = 2/\sqrt{27}$	0.500 = 1/2	0.500 = 1/2
0.1	0.702	0.556	0.680	0.391	0.524	0.524
0.2	0.743	0.586	0.694	0.398	0.556	0.551
0.3	0.790	0.621	0.710	0.405	0.598	0.581
0.4	0.844	0.663	0.729	0.413	0.652	0.614
0.5	0.900 = 9/10	0.711 = 32/45	0.750 = 3/4	0.422 = 27/64	0.724 $= (5+\sqrt{5})/10$	0.655 $= (3+\sqrt{5})/8$
0.6	0.951	0.768	0.775	0.432	0.814	0.704
0.7	0.986	0.830	0.806	0.444	0.912	0.766
0.8	0.999	0.893	0.845	0.458	0.984	0.843
0.9	1.000	0.951	0.898	0.475	1.000	0.926
1.0	{1.000 = 1}	1.000 = 1	{1.000 = 1}	0.500 = 1/2	{1.000 = 1}	1.000 = 1

constant-linear		quadratic		linear		constant	
α	υ	α	υ	α	υ	α	υ
1/2	1/2	1/2	1/4	1	1	0	1

Table 2
Information matrices and their eigenvalues of symmetric three-point designs ξ_α, $\alpha \in (0,1)$, for four parameter systems. The last row gives the expression whose zero determines the optimal weight $\alpha(p)$

	constant-linear-quadratic	linear-quadratic	constant-quadratic	constant-linear
Information matrix	$\begin{bmatrix} 1 & 0 & \alpha \\ 0 & \alpha & 0 \\ \alpha & 0 & \alpha \end{bmatrix} = M_\alpha$	$\begin{pmatrix} \alpha & 0 \\ 0 & \alpha(1-\alpha) \end{pmatrix}$	$\begin{pmatrix} 1 & \alpha \\ \alpha & \alpha \end{pmatrix}$	$\begin{pmatrix} 1-\alpha & 0 \\ 0 & \alpha \end{pmatrix}$
Eigenvalues	λ, α, Λ	$\alpha, \alpha(1-\alpha)$	λ, Λ	$\alpha, 1-\alpha$
$\alpha(p)$ is zero of	$\lambda'\lambda^{p-1}+\alpha^{p-1}+\Lambda'\Lambda^{p-1}$	$1+(1-2\alpha)(1-\alpha)^{p-1}$	$\lambda'\lambda^{p-1}+\Lambda'\Lambda^{p-1}$	$\alpha-\frac{1}{2}$

$$\text{Maximize } f_p(\alpha)=j_p(J(M_\alpha)), \text{ subject to } \alpha \in [0, 1].$$

Omitting the discussion of the boundary weights $\alpha=0$ and $\alpha=1$ we assume $\alpha \in (0, 1)$. Set

$$\lambda=\tfrac{1}{2}(1+\alpha)-w, \qquad \Lambda=\tfrac{1}{2}(1+\alpha)+w, \qquad w=(\tfrac{1}{4}(1-\alpha)^2+\alpha^2)^{1/2}.$$

These quantities satisfy the relation $0<\lambda<\alpha<1<\Lambda$, and appear as eigenvalues of the information matrices for subsets which contain more than one parameter, as displayed in Table 2. All eigenvalues λ, α, Λ of the moment matrices M_α are increasing for $\alpha \in (0, 0.4)$, whence an optimal weight α must satisfy $\alpha \geq 0.4$.

It is now straightforward to compute, as a function of α, the derivatives λ' and Λ'. For $-\infty<p<1$ the function f_p is strictly concave. Hence the optimal weight $\alpha(p)$ is the unique zero of the derivative f_p', or equivalently, of the expression given in the last row of Table 2. Newton–Raphson or bisection iteration was used for numerical computation. The optimal information value then is found from the formula $v(p)=f_p(\alpha(p))$. The graph of each curve is based on 500 points. More details are given in Preitschopf and Pukelsheim (1985).

4. Other criteria based on generalized means

Alternatively we may maximize the mean of the diagonal elements of the information matrix for the parameters of interest. In case of the constant-linear-quadratic, and the constant-quadratic parameters with information matrices as given in Table 2 the diagonal elements are maximized when $\alpha=1$. It is pleasing that the optimal weight $\alpha=1$ does not depend on the criterion parameter p, it is distracting that the parameters of interest fail to be identifiable.

As an alternative to maximizing information we may wish to minimize dispersion by using a *convex* mean of order r, i.e. $1 \leq r \leq \infty$. However, r-mean minimization of the eigenvalues of the dispersion matrices $(J(M_\alpha))^{-1}$ coincides with p-mean maximization, $p=-r$, of the eigenvalues of the information matrices $J(M_\alpha)$. Hence we are led back to a subset of solutions as presented in Section 3.

Finally we mention minimization of the diagonal elements of the dispersion matrices for the parameters of interest when the criterion is a convex r-mean. For instance, the case $r=\infty$, i.e. minimizing the maximal diagonal element, is listed as criterion (d) in Kiefer (1960, p. 383). For the constant-linear-quadratic parameters the optimal weight is found to be $\alpha=\frac{1}{2}$ and does not depend on the criterion parameter r. Now let $\alpha(p)$ be the optimal weight for the p-mean of the information eigenvalues for the linear-quadratic parameters as discussed in Section 3. For this parameter subset the weight minimizing r-means of the diagonal elements of the dispersion matrix $(J(M_\alpha))^{-1}$ then is found to be $\alpha(-r)$, while for the constant-quadratic parameters it turns out to be $1-\alpha(-r)$ with the optimal weight function $\alpha(p)$ of the linear-quadratic case. Notice that in the latter two cases the monotonicity behaviour of the optimal weight function is reversed.

In summary this discussion would seem to suggest that the Kiefer criteria, of maximizing concave means of information eigenvalues, are those which rightly carry the greatest interest.

References

Z. Galil and J.C. Kiefer (1977). Comparison of rotatable designs for regression on balls, I (quadratic). *J. Statist. Plann. Inference* 1, 27–40.

A. Giovagnoli, F. Pukelsheim and H.P. Wynn (1987). Group invariant orderings and experimental designs. *J. Statist. Plann. Inference* 17, to appear.

J.C. Kiefer (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* 21, 272–319.

J.C. Kiefer (1960). Optimum experimental designs V, with applications to systematic and rotatable designs. *Proc. 4th Berkeley Symp. Math. Statist. Probab.* 1, 381–405.

J.C. Kiefer (1975). Construction and optimality of generalized Youden designs. In: J.N. Srivastava, Ed., *A Survey of Statistical Design and Linear Models* 333–353.

F. Preitschopf and F. Pukelsheim (1985). Optimal designs for quadratic regression. Preprint Nr. 92, Institut für Mathematik, Universität Augsburg.

F. Pukelsheim (1980). On linear regression designs which maximize information, *J. Statist. Plann. Inference* 4, 339–364.

W.J. Studden (1980). D_s-optimal designs for polynomial regression using continued fractions. *Ann. Statist.* 8, 1132–1141.