NOTES

TWO MORE CRITERIA EQUIVALENT TO D-OPTIMALITY OF DESIGNS[1]

By J. Kiefer

Cornell University

0. Summary. Two minimax "regret" criteria for global optimality of designs in terms of estimation of the entire regression function are shown to be equivalent to minimizing the generalized variance. (The equivalence for the minimax criterion *without* modification was proved by Kiefer and Wolfowitz [9].) A consequent algorithm which is helpful in computing optimum designs is given.

1. Introduction. Let f be a column k-vector of continuous functions f_i on a compact space X. The expected value of an observation Y_x corresponding to the value x of the independent variable is $\theta' f(x) = \sum_1^k \theta_i f_i(x)$, where the vector θ of regression coefficients is unknown. A design is a (discrete) probability measure ξ on X, where for each x the value $\xi(x)$ denotes the proportion of observations taken at x. The observations are assumed uncorrelated and have common variance σ^2. Let $M(\xi)$ denote the $k \times k$ matrix

$$M(\xi) = \int_X f(x)f(x)'\xi(dx).$$

If N observations are taken according to ξ, then $N\sigma^{-2}M(\xi)$ is the information matrix of this design; if nonsingular, this is the inverse of the covariance matrix of best linear estimators of θ. In the approximate theory wherein Elfving [3], Chernoff [1], Kiefer and Wolfowitz [8], and others have successfully characterized designs which are optimum in various senses for such regression problems, we do not restrict $N\xi$ to be integral-valued, but instead allow ξ to be an arbitrary (discrete) probability measure. This permits useful results to be obtained, and a single optimum ξ can be used to yield, for each N, an actual design which is optimum to within order $1/N$. We shall be working in the approximate theory in this note.

A design ξ^* is said to be D-*optimum* (minimizes the generalized variance of best linear estimators of θ) if

$$\det M(\xi^*) = \max_\xi \det M(\xi). \tag{1}$$

If $M(\xi)$ is nonsingular, write $d(x, \xi) = f(x)'M^{-1}(\xi)f(x)$. Then $N^{-1}\sigma^2 d(x, \xi)$ is the variance of the (best linear) estimated regression at x. We also define $d(x, \xi)$ to conform with this if x is such that $\theta' f(x)$ is estimable but $M(\xi)$ is

Received June 27, 1961; revised December 4, 1961.

[1] Research sponsored by the Office of Naval Research.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

singular, and put $d(x, \xi) = \infty$ for those x for which $\theta' f(x)$ is not estimable under ξ. A design ξ^* is said to be G-*optimum* (globally optimum) if

$$\max_x d(x, \xi^*) = \min_\xi \max_x d(x, \xi). \tag{2}$$

It is not hard to show that $\max_x d(x, \xi) \geqq k$ for every ξ. It was proved by Kiefer and Wolfowitz [9] that ξ^* is D-optimum if and only if it is G-optimum, and if and only if

$$\max_x d(x, \xi^*) = k. \tag{3}$$

This equivalence was extended by the author to the case where we are interested in s out of the k parameters, and it has proved useful in the computation of optimum designs in various settings [6], [7].

Let

$$d(x) = \min_\xi d(x, \xi). \tag{4}$$

This quantity is proportional to the variance of the best linear estimator of $\theta' f(x)$ for the design which is best for that x. An optimality criterion which is analogous to the minimax regret criterion of decision theory was suggested to the author by Professor Herman Rubin: ξ^* will be said to be MR-*optimum* (optimum in the sense of multiplicative regret) if

$$\max_x [d(x, \xi^*)/d(x)] = \min_\xi \max_x [d(x, \xi)/d(x)]. \tag{5}$$

A similar criterion in terms of additive regret is to call ξ^* AR-*optimum* if

$$\max_x [d(x, \xi^*) - d(x)] = \min_\xi \max_x [d(x, \xi) - d(x)]. \tag{6}$$

The purpose of this note is to give a very simple proof of the following.

THEOREM. *ξ^* is* D- (*and* G-) *optimum if and only if it is* MR-*optimum, and if and only if it is* AR-*optimum.*

This result incidentally yields a computational algorithm which is discussed in Section 3.

We make three brief remarks before proceeding to the proof. Firstly, the analogous results in the *exact* theory in the symmetric settings where symmetric block designs (Latin squares, BIBD's, etc.) are customarily used (the f_i taking on values 0 and 1) follow at once from the constancy of $d(x)$ and the results of [4]; this is not of much practical interest, since estimated regression is of less concern than estimated contrasts of treatment or variety effects in such settings, where block effects, row effects, etc., are of no interest. Secondly, the function-space interpretation of the equivalence of (1), (2) and (3) which was mentioned in [9] has an addition in terms of (5) and (6) which the reader will find it no difficulty to state. Thirdly, since the analogue of $d(x, \xi)$ in the case $s < k$ (see [6]) has an interpretation which depends on ξ (it is not merely proportional to the variance of the best linear estimator of $\sum_1^s \theta_i f_i(x)$), there is no obvious meaningful extension of the theorem of this note to that case.

2. Proof of the theorem. We first prove that (3) implies (5). Let ξ_x assign measure one to the point x. The regression $\theta' f(x)$ is clearly estimable at x if the design ξ_x is used, with N[variance of estimated regression at x under $\xi_x]/\sigma^2 = 1$. Hence,

$$d(x) \leqq 1. \tag{7}$$

Suppose ξ^* is D-optimum. It follows from (7) and the sentence following (2) that, for every design ξ,

$$\max_x [d(x, \xi)/d(x)] \geqq k. \tag{8}$$

We shall show that

$$\max_x [d(x, \xi^*)/d(x)] = k, \tag{9}$$

which will thus prove that ξ^* is MR-optimum.

If (9) is not satisfied, then there is a design ξ' and a value x' such that

$$d(x', \xi^*)/d(x', \xi') > k. \tag{10}$$

By replacing ξ' by $\epsilon\xi^* + (1 - \epsilon)\xi'$ with ϵ small, we can assume (10) is satisfied by a nonsingular ξ'. We can assume $M(\xi^*)$ to be the identity and $M(\xi')$ to be diagonal with diagonal elements d_i, since otherwise these matrices can be so diagonalized by a linear transformation without affecting the proof. According to (10),

$$1/k > \sum_{i=1}^{k} \left\{ f_i^2(x') \Big/ \sum_{j=1}^{k} f_j^2(x') \right\} d_i^{-1}. \tag{11}$$

Hence, at least one d_i^{-1} is $<k^{-1}$, say $d_1 > k$. But then

$$\begin{aligned} &\frac{\partial}{\partial\alpha} \log \det M(\alpha\xi' + (1 - \alpha)\xi^*) \Big|_{\alpha=0+} \\ &= \frac{\partial}{\partial\alpha} \sum_i \log (1 - \alpha + \alpha d_i) \Big|_{\alpha=0} \\ &= \sum_i (d_i - 1) > 0, \end{aligned} \tag{12}$$

contradicting the fact (1) that det $M(\xi)$ is a maximum for $\xi = \xi^*$.

The converse is trivial: if ξ^* is MR-optimum, it must satisfy (9), since any D-optimum design satisfies (9); but then, by (7), we have $\max_x d(x, \xi^*) \leqq k$, so that ξ^* is D-optimum.

To prove the equivalence of D-optimality to AR-optimality, it is only necessary, in the previous two paragraphs, to replace every appearance of (8) by

$$\max_x [d(x, \xi) - d(x)] \geqq k - 1, \tag{8'}$$

of (9) by

$$\max_x [d(x, \xi^*) - d(x)] = k - 1, \tag{9'}$$

of (10) by

$$d(x', \xi^*) - d(x', \xi') > k - 1, \tag{10'}$$

and of (11) by

$$k^{-1} \geqq 1 - \frac{k-1}{\sum_j f_j^2(x')} > \sum_i \{f_i^2(x')/\sum_j f_j^2(x')\} d_i^{-1}, \tag{11'}$$

the first half of (11′) being a consequence of (3) and the second half following from (10′). The rest of the proof then reads as before.

3. A computational algorithm. It follows from (7), (9), and the fact [9] that any D-optimum ξ^* assigns measure one to a set of values x for which $d(x, \xi^*) = k$, that any such ξ^* assigns measure one to the set

$$B = \{x : d(x) = 1\}.$$

Thus, if B is much smaller than X, the characterization of B can be of aid in computing a D-optimum ξ^*. In the present section we describe a characterization of B in terms of a family of related Chebyshev approximation problems, bringing out a relationship between the problems of estimating one and k parameters.

Before obtaining this characterization we remark that, if ξ^* is admissible in the sense of Ehrenfeld [2] (see also [5]), then, since multiples of $\theta' f(x)$ are the only estimable linear parametric functions when ξ_x is used, we must have $d(x) = 1$ and thus $x \varepsilon B$. This indicates that B need not be much of a reduction from X; for example, in the case of polynomial regression in one variable, the admissible ξ's were characterized in [5] and include all ξ_x's, so that $B = X$ in that case. However, it is easy to give other examples where B is smaller than X. Moreover, one can sometimes find a proper subset B' of B which must support an optimum ξ^*: if x in $B - B'$ implies that there is a ξ' on B' which is at least as good as ξ_x (that is, such that $M(\xi') - M(\xi_x)$ is nonnegative-definite), then this is the case.

We now describe an algorithm for computing $d(x_0)$. Suppose, without loss of generality, that $f_1(x_0) \neq 0$. (Any x for which all $f_i(x) = 0$ can be deleted from X.) Let $\phi_1 = \theta' f(x_0)$ and $\phi_i = \theta_i$ for $i \geqq 2$, and let $g_1(x) = f_1(x)/f_1(x_0)$ and $g_i(x) = f_i(x) - f_i(x_0) g_1(x)$ for $i \geqq 2$. Then

$$\sum \phi_i g_i(x) = \sum \theta_i f_i(x),$$

and the problem of estimating $\theta' f(x_0)$ when the regression is $\theta' f(x)$ is the same as that of estimating ϕ_1 when the regression is $\phi' g(x)$. The latter problem was first attacked in [3], and the following algorithm was given in [8]: Let $c^* = (c_2^*, \cdots, c_k^*)$ yield a best linear Chebyshev approximation of g_1 by $g_2, \cdots, g_k$; that is, the quantity

$$m(c) = \max_{x \varepsilon X} |g_1(x) - \sum_2^k c_i g_i(x)|$$

is minimized by the choice $c = c^*$. A design which minizes $d(x_0)$ can then be obtained easily from c^* in a manner described in [8]; for our present considerations, we need only mention that $d(x_0) = [m(c^*)]^{-2}$, which can be used to tell us whether or not $x_0 \varepsilon B$.

Finally, we remark that the Chebyshev approximation problem just described in terms of the g_i's can be rewritten as a "modified Chebyshev problem" in terms of the original f_i's, namely, to minimize

$$\max_x |[1 + \sum_2^k c_i f_i(x_0)] f_1(x)/f_1(x_0) - \sum_2^k c_i f_i(x)|.$$

For computational purposes, it is often convenient to solve this problem by first solving the restricted Chebyshev problem of minimizing

$$\max_x |f_1(x)/f_1(x_0) - r^{-1} \sum_2^k c_i f_i(x)|$$

subject to $\sum_2^k c_i f_i(x_0) = r - 1$, then multiplying the resulting minimum by r and minimizing with respect to r.

REFERENCES

[1] CHERNOFF, H. (1953). Locally optimum designs for estimating parameters. *Ann. Math. Statist.* **24** 586–602.

[2] EHRENFELD, S. (1955). Complete class theorems in experimental designs. *Proc. Third Berkeley Symp. Math. Stat. Prob.* **1** 57–67. Univ. of California Press.

[3] ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 225–262.

[4] KIEFER, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *Ann. Math. Statist.* **29** 675–699.

[5] KIEFER, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc., Ser. B* **21** 272–319.

[6] KIEFER, J. (1961). Optimum designs in regression problems, II. *Ann. Math. Statist.* **32** 298–325.

[7] KIEFER, J. (1961). Optimum experimental designs, *V*, with applications to rotatable and systematic designs. *Proc. Fourth Berkeley Symp. Math. Stat. Prob.* **1** 381–405. Univ. of California Press.

[8] KIEFER J. and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

[9] KIEFER J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Can. J. Math.* **12** 363–366.

A CONTOUR-INTEGRAL DERIVATION OF THE NON-CENTRAL CHI-SQUARE DISTRIBUTION

BY FRANK MCNOLTY

Lockheed Missiles and Space Division, Palo Alto

The brief discussion which follows presents a contour-integral derivation of the non-central chi-square distribution. Although this distribution is well

Received March 29, 1961; revised November 4, 1961.