THE EQUIVALENCE OF TWO EXTREMUM PROBLEMS

J. KIEFER AND J. WOLFOWITZ

1. Introduction. Let $f_1, \ldots, f_k$ be linearly independent real functions on a space X, such that the range R of $(f_1, \ldots, f_k)$ is a compact set in k-dimensional Euclidean space. (This will happen, for example, if the f_i are continuous and X is a compact topological space.) Let S be any Borel field of subsets of X which includes X and all sets which consist of a finite number of points, and let $C = \{\xi\}$ be any class of probability measures on S which includes all probability measures with finite support (that is, which assign probability one to a set consisting of a finite number of points), and which are such that

$$m_{ij}(\xi) = \int_X f_i(x)f_j(x)\xi(dx) \qquad i, j = 1, \ldots, k$$

is defined. In all that follows we consider only probability measures ξ which are in C. Write $M(\xi)$ for the $k \times k$ matrix $||m_{ij}(\xi)||$. When $M(\xi)$ is non-singular, write $[M(\xi)]^{-1} = ||m^{ij}||$. (We shall not always exhibit dependence on ξ.) Letting $f(x)$ denote the column vector with components $f_i(x)$, and letting primes denote transposes, we define

$$d(x; \xi) = f(x)'[M(\xi)]^{-1}f(x)$$

whenever $M(\xi)$ is non-singular.

We consider two extremum problems. The first is to choose ξ so that

(1) $\qquad\qquad\qquad \xi \text{ maximizes } \det M(\xi).$

The second is to choose ξ so that

(2) $\qquad\qquad\qquad \xi \text{ minimizes } \max_x d(x; \xi).$

We also note that the integral with respect to ξ of $d(x; \xi)$ is k; hence, $\max_x d(x; \xi) \geqslant k$, and thus a sufficient condition for ξ to satisfy (2) is

(3) $\qquad\qquad\qquad \max_x d(x; \xi) = k.$

The result of this note is that (1), (2), and (3) are equivalent. This result, which seems to have interest *per se*, also strengthens and extends results of the authors **(1)** on the optimum design of regression experiments. A brief description of the connection with the design of such experiments is given below. The proof of the theorem is elementary and brief.

2. The theorem. For every ξ consider $M(\xi)$ as a point in Euclidean k^2-space, let T be the totality of such points for all ξ in C, and let $\bar{T}$ be the convex closure of T. It is clear that every extreme point of $\bar{T}$ can be achieved by a ξ which assigns probability one to a single point. Since C contains every ξ with finite support, it follows that $T = \bar{T}$. The class C need not, of course, be convex. However, since our argument will be concerned only with the $M(\xi)$, we may argue below as if C were convex. Thus, if ξ_1 and ξ_2 are in C and

$$\frac{\xi_1 + \xi_2}{2}$$

is not, we may still discuss

$$M\left(\frac{\xi_1 + \xi_2}{2}\right),$$

because there exists a ξ in C with finite support, say ξ_3, such that

$$M(\xi_3) = M\left(\frac{\xi_1 + \xi_2}{2}\right).$$

Moreover, if $H - 1$ is the dimension of the linear space spanned by the functions $f_i f_j$, $i \leqslant j$, any $M(\xi)$ is equal to an $M(\xi')$ where the support of ξ' consists of at most H points. This can often be impoved, as in the case where X is the unit interval and $f_i(x) = x^{i-1}$.

Call a subset D of C *linear* if the following condition holds: For every α, $0 \leqslant \alpha \leqslant 1$, and every pair ξ_1, ξ_2 in D, $\alpha\xi_1 + (1 - \alpha)\xi_2$ is in D whenever it is in C. Thus, if C is convex, D is also convex.

We shall prove the following:

THEOREM. *Conditions (1), (2), and (3) are equivalent. The set B of all ξ satisfying these conditions is linear, and $M(\xi)$ is the same for all ξ in B.*

This result has a function space corollary which may be of interest. Suppose ξ satisfies (3) and that Q is a real $k \times k$ matrix such that $QM(\xi)Q'$ is the identity. Then $g = Qf$ is a vector of orthonormal functions with respect to ξ, and $g(x)'g(x) = d(x; \xi)$. Thus we have

COROLLARY. *If $f_1, \ldots, f_k$ are linearly independent, continuous, real functions on a compact space X, then there is a probability measure ξ on X and a linear transformation $g_i = \sum_j a_{ij}f_j$ such that $g_1, \ldots, g_k$ are orthonormal with respect to ξ and*

$$\max_x \sum_{i=1}^k g_i^2(x) = k.$$

The set of all such ξ is the set B of the theorem.

Proof of the theorem. We shall say that ξ is a *local solution* of (1) if $\det M(\xi) > 0$ and if, for every ξ',

(4) $$\frac{\partial}{\partial \alpha} \log \det M([1-\alpha]\xi + \alpha\xi')|_{\alpha=0+} \leqslant 0.$$

Now, if $\det M(\xi) > 0$, A is such that $AM(\xi)A'$ is the identity, and $AM(\xi')A'$ is diagonal with diagonal elements b_i, then $\det M([1-\alpha]\xi + \alpha\xi') = \det A^{-2}\Pi_i[1-\alpha+\alpha b_i]$, from which we easily compute that $-\log\det M([1-\alpha]\xi + \alpha\xi')$ is convex in $\alpha (0 < \alpha < 1)$ and is strictly convex unless all $b_i = 1$ (that is, unless $M(\xi) = M(\xi')$). Hence, if $\det M(\xi') > \det M(\xi)$, equation (4) cannot hold for that ξ'. We conclude that local solutions of (1) are actual solutions of (1), and of course the converse is true. Moreover, if $\det M(\xi) = \det M(\xi') = h > 0$, we have $\det M(\xi/2 + \xi'/2) > h$ unless $M(\xi) = M(\xi')$, so that ξ and ξ' cannot both satisfy (1) unless $M(\xi) = M(\xi')$. It follows from this and the linearity in ξ of $M(\xi)$ that, if ξ and ξ' both satisfy (1), then so does $\alpha\xi + (1-\alpha)\xi'$, whenever it is in C.

It now suffices to prove that $\det M(\xi) > 0$ and ξ satisfies (4) for all ξ', if and only if ξ satisfies (2), and only if it satisfies (3). First suppose ξ satisfies (4) and that $\det M(\xi) > 0$. Performing the differentiation in (4), and denoting by M_{ij} the cofactor of m_{ij}, we have

(5) $$0 \geqslant [\det M(\xi)]^{-1} \sum_{i,j} \frac{\partial \det M}{\partial m_{ij}} \frac{\partial m_{ij}([1-\alpha]\xi + \alpha\xi')}{\partial \alpha}\bigg|_{\alpha=0}$$

$$= [\det M(\xi)]^{-1} \sum_{i,j} \left(\frac{\partial}{\partial m_{ij}} \sum_q m_{iq}M_{iq}\right)[m_{ij}(\xi') - m_{ij}(\xi)]$$

$$= [\det M(\xi)]^{-1} \sum_{i,j} M_{ij}(\xi)[m_{ij}(\xi') - m_{ij}(\xi)] = \sum_{i,j} m^{ij}(\xi)m_{ij}(\xi') - k.$$

Letting ξ' give measure one to the point x, we obtain

(6) $$[f(x)]'M(\xi)^{-1}f(x) \leqslant k$$

for all x. Thus, (3) is satisfied and, as we have remarked, this implies (2).

Finally, if (2) is satisfied, we must have (6) for all x, since we have just seen that there always exist ξ's satisfying (3). Hence, for any ξ' with finite support, we obtain $\sum_{i,j}m^{ij}(\xi)m_{ij}(\xi') \leqslant k$. Hence this inequality is valid for all ξ', and (5) is satisfied. This completes the proof of the theorem.

3. Extensions and applications. We remark that it is easy to see that, if R is bounded but not compact, and if $\{\xi_i\}$ is a sequence of measures on S, then $\lim_i\det M(\xi_i)$ is a maximum if and only if $\lim_i \sup_x d(x;\xi_i)$ is a minimum, and if and only if $\lim_i \sup_x d(x;\xi_i) = k$. Similarly, the first part of the corollary holds with the replacement $\sup_x \sum_i g_i{}^2(x) < k + \epsilon$, for any $\epsilon > 0$.

We now describe briefly the statistical applications of the results. An integer N is given, and the statistician must choose N points $x_1, \ldots, x_N$ (not necessarily distinct) corresponding to which he obtains observations on uncorrelated random variables Y_i $(1 \leqslant i \leqslant N)$ with common variance σ^2 (perhaps unknown) and with expectation $\sum_{j=1}^{k}\theta_j f_j(x_i)$, where the θ_j are unknown

real parameters. If $\xi(x)$ denotes the proportion of x_i's which are equal to x, we find that the covariance matrix of best linear estimators of $\theta_1, \ldots, \theta_k$ is $N^{-1}\sigma^2[M(\xi)]^{-1}$. The function ξ is called the experiment or the experimental design. A criterion often adopted for choosing a design is to minimize the determinant of the above covariance matrix (the "generalized variance"). Another possible criterion is to minimize the maximum over x of the variance $N^{-1}\sigma^2 d(x;\xi)$ of the "best linear estimator," given ξ, of the "regression function" $\sum_j \theta_j f_j(x)$. If we consider not merely the class C_N of probability measures ξ which take on only integral multiples of N^{-1} as values, but rather *all* probability measures ξ in C, then our result is that the two optimality criteria are equivalent. Moreover, for any ξ with support on H points which satisfies (1), (2), and (3), there is clearly a ξ' in C_N which achieves (1), (2), and (3) to within a multiplicative factor $1 + 0(N^{-1})$, and is easy to write down from ξ. Since the *exactly* optimum designs are often difficult to obtain, depend on N, and differ for the two criteria, we see the practical importance of our considerations.

It is very helpful to use the interplay of the two criteria (1) and (2) in obtaining a solution. For example, one can sometimes guess that a solution exists which is a member of a class of ξ which depend on several parameters. One may use (1) as the more convenient initial approach, maximize det $M(\xi)$ over the parametric class, and then verify whether the maximum just obtained is indeed a maximum over *all* ξ (which may be difficult in terms of (1)) by verifying (3). It is useful to note that, if ξ has a set consisting of k points as its support, then it gives equal measure to each of these points. (This is part of Theorem 5 of **(1)**.) Examples which make use of such methods will appear elsewhere, as will generalizations such as one concerned with the minimization of the determinant of a principal minor of $M(\xi)^{-1}$.

REFERENCE

1. J. Kiefer and J. Wolfowitz, *Optimum designs in regression problems*, Ann. Math. Stat., *30* (1959).

Cornell University