AN EXTREMUM RESULT

J. KIEFER

1. Introduction. The main object of this paper is to prove the following:

THEOREM.† *Let $f_1, \ldots, f_k$ be linearly independent continuous functions on a compact space $\mathfrak{X}$. Then for $1 \leqslant s \leqslant k$ there exist real numbers a_{ij}, $1 \leqslant i \leqslant s$, $1 \leqslant j \leqslant k$, with $\{a_{ij}, 1 \leqslant i, j \leqslant s\}$ non-singular, and a discrete probability measure ξ^* on $\mathfrak{X}$, such that*

(a) *the functions $g_i = \sum_{j=1}^{k} a_{ij} f_j$ $1 \leqslant i \leqslant s$, are orthonormal ($\xi^*$) and are orthogonal ($\xi^*$) to the f_j for $s < j \leqslant k$;*

(b) $$\max_{x \in \mathfrak{X}} \sum_1^s f_i^2(x) = \int_{\mathfrak{X}} \sum_1^s f_i^2(x) \xi^*(dx) = s.$$

The result in the case $s = k$ was first proved in **(2)**. The result when $s < k$, which because of the orthogonality condition of (a) is more general than that when $s = k$, was proved in **(1)** under a restriction which will be discussed in § 3. The present proof does not require this *ad hoc* restriction, and is more direct in approach than the method of **(2)** (although involving as much technical detail as the latter in the case when the latter applies). The latter proof involved showing the equivalence of two extremum problems encountered in the theory of optimum statistical designs, under the restriction mentioned. The analogous equivalence result when the restriction is not satisfied is more difficult to state and to prove, and cannot be used to obtain the result of the present paper without proving an additional fact, discussed in § 3, which the methods of **(1)** do not seem to yield. On the other hand, the present result implies that additional fact in the design setting, and yields a neater proof of a major part of the equivalence result when the mentioned restriction is not necessarily satisfied. With or without the restriction, the remainder of the equivalence theorem can then be given a short proof.

The idea of the present proof is to reduce the result to one of the ontoness of a certain natural mapping induced by the problem. This ontoness, which seems intuitively plausible, is not so easy to prove, and the author is indebted

Received July 31, 1961. Research for this paper was sponsored by the Office of Naval Research of the United States.

†The same conclusion obviously holds if $\dim\{(f_1(x), \ldots, f_k(x)); x \in \mathfrak{X}\} = s + \dim\{(f_{s+1}(x), \ldots, f_k(x)); x \in \mathfrak{X}\}$, and the necessary modification in the statement of the theorem in other cases is obvious. There is a choice of ξ^* whose support consists of no more than $s(2k-s+1)/2$ points **(1)**. The a_{ij}, $i > j$, can be taken to be zero.

to his colleague Professor Namioka for supplying a proof of a much more general theorem **(3)** which yields this ontoness (as well as other interesting results), and for many helpful discussions.

2. Proof of the theorem. The proof will be divided into three parts for convenience.

I. Let $\mathfrak{C}$ be the set of vectors $c = (c_{ij}, 1 \leqslant i \leqslant s, i < j \leqslant k)$. Let $\Lambda = \{(\lambda_1, \ldots, \lambda_s) : \lambda_i \geqslant 0, \sum \lambda_i = 1\}$. For $\lambda \in \Lambda$, write

$$K_\lambda(x, c) = \sum_{i=1}^{s} \lambda_i \Big[f_i(x) - \sum_{j>i} c_{ij} f_j(x) \Big]^2.$$

We consider the game $\{K_\lambda, \mathfrak{X}, \mathfrak{C}\}$ with $\mathfrak{X}$ and $\mathfrak{C}$ the spaces of pure strategies and K_λ the payoff function. Writing $K_\lambda(\xi, c) = \int K_\lambda(x, c)\xi(dx)$, we as usual define c^* to be minimax if $\max_x K_\lambda(x, c^*) = \min_c \max_x K_\lambda(x, c)$, and ξ^* to be maximin if $\min_c K_\lambda(\xi^*, c) = \max_\xi \min_c K_\lambda(\xi, c)$. Since K_λ is convex in c, it suffices to consider pure strategies c for player 2, but we must allow mixed strategies ξ for player 1. The range of $\{f_j, 1 \leqslant j \leqslant k\}$ is a compact Euclidean set which could actually be regarded as $\mathfrak{X}$, the associated class Ξ of Borel measures being weakly compact, which will be used below. There are never any measure-theoretic difficulties, and in fact any ξ can be replaced by a ξ' with finite support and such that $K_\lambda(\xi, c) = K_\lambda(\xi', c)$ for all λ and c.

Let $\mathfrak{C}_N$ be the subset of vectors c satisfying $\sum_{j=i+1}^{k} c_{ij}^2 \leqslant N$, $1 \leqslant i \leqslant s$, and let $\mathfrak{C}_i' = \{(c_{i,i+1}, \ldots, c_{i,k}) : \sum_{j=i+1}^{k} c_{ij}^2 = 1\}$. For each i, $1 \leqslant i \leqslant s$, the quantity

$$b_i(\xi) = \min_{\{c_{ij}\} \in \mathfrak{C}_i'} \int \Big[\sum_{j>i} c_{ij} f_j(x) \Big]^2 \xi(dx)$$

is positive for some $\xi \in \Xi$ because the f_i are linearly independent. Hence (averaging the ξ's for different i's) there is a $\xi' \in \Xi$ and an $\epsilon > 0$ such that $b_i(\xi') > \epsilon$ for $1 \leqslant i \leqslant s$. Thus, there is a constant d such that

$$\min_{c \notin \mathfrak{C}_N} K_\lambda(\xi', c) > N\epsilon - d$$

for all $\lambda \in \Lambda$. It follows that there is a value N' of N such that, for any $c \notin \mathfrak{C}_{N'}$,

$$\max_\xi K_\lambda(\xi, c) > \min_{c'} \max_\xi K_\lambda(\xi, c'),$$

for all λ.

Hence, c^*_λ is minimax for the game $\{K_\lambda, \mathfrak{X}, \mathfrak{C}\}$ if and only if it is minimax for $\{K_\lambda, \mathfrak{X}, \mathfrak{C}_{N'}\}$. The latter game, having compact strategy spaces and continuous payoff, is determined. For $N > N'$, if ξ^*_N is maximin for the game $\{K_\lambda, \mathfrak{X}, \mathfrak{C}_N\}$ and c^*_λ is minimax (for all $N > N'$), and if $\xi_j = [(j-1)\xi^*_j + \xi']/j$ and $N_j > N'$ is such that $K_\lambda(\xi_j, c^*_\lambda) < K_\lambda(\xi_j, c)$ for all $c \notin \mathfrak{C}_{N_j}$, we have (since $K_\lambda(\xi^*_j, c^*_\lambda) = \max_\xi K(\xi, c^*_\lambda)$ by the determinateness),

$$\begin{aligned}\sup_\xi \inf_{c\in\mathfrak{C}} K_\lambda(\xi, c) &\geqslant \inf_{c\in\mathfrak{C}} K_\lambda(\xi_j, c)\\ &= \inf_{c\in\mathfrak{C}_{N_j}} K_\lambda(\xi_j, c) \geqslant \left(1-\frac{1}{j}\right) \inf_{c\in\mathfrak{C}_{N_j}} K_\lambda(\xi_j^*, c)\\ &= \left(1-\frac{1}{j}\right) K_\lambda(\xi_j^*, c_\lambda^*) = \left(1-\frac{1}{j}\right)\sup_\xi K_\lambda(\xi, c_\lambda^*)\\ &\geqslant \left(1-\frac{1}{j}\right)\inf_{c\in\mathfrak{C}} \sup_\xi K_\lambda(\xi, c).\end{aligned}$$

Letting $j \to \infty$, we see that the game $\{K_\lambda, \mathfrak{X}, \mathfrak{C}\}$ is also determined.

Moreover, the set C_λ (say) of (pure) minimax strategies for this last game is a subset of $\mathfrak{C}_{N'}$, for all λ. Since $\max_x K_\lambda(x, c)$ is convex in c (being a maximum of convex functions), we have that C_λ is convex. Let Ξ_λ be the (mixed) maximin strategies for this game. Since $K_\lambda(\xi, c)$ is linear in ξ, we similarly have that Ξ_λ is convex.

Let $\{\lambda_n\}$ be a sequence in Λ, converging to λ_0. If $c^*{}_{\lambda_n}$ and $\xi^*{}_{\lambda_n}$ are any elements of C_{λ_n} and Ξ_{λ_n}, we can, by the compactness of $\mathfrak{C}_{N'}$ and Ξ, select a subsequence $\{n_m\}$ of n such that $\{c^*{}_{\lambda_{n_m}}\}$ and $\{\xi^*{}_{\lambda_{r_m}}\}$ converge to limits c_0 and ξ_0 in $\mathfrak{C}_{N'}$ and Ξ. Since $\min_c K_\lambda(\xi, c)$ and $\max_x K_\lambda(x, c)$ are clearly continuous in (λ, ξ) and (λ, c), respectively (because the f_i are continuous), we see at once that $c_0 \in C_{\lambda_0}$ and $\xi_0 \in \Xi_{\lambda_0}$. We conclude that, if in the space $\Lambda \times \mathfrak{C}_{N'} \times \Xi$ we let $V_\lambda = (\{\lambda\} \times C_\lambda \times \Xi_\lambda)$ and

$$G = \cup_{\lambda\in\Lambda} V_\lambda,$$

then G is closed and each section V_λ is a compact, convex subset of the compact space $\mathfrak{C}_{N'} \times \Xi$.

II. Let k_λ denote the value $\min_c \max_x K_\lambda(x, c) = \max_\xi \min_c K_\lambda(\xi, c)$ of the above game. Since

$$\min_c K_\lambda(\xi, c) \geqslant \min_i \min_{\{c_{ij}\}} \int \left[f_i(x) - \sum_{j>i} c_{ij} f_j(x)\right]^2 d\xi$$

and the inner minimum is the square of the $L_2^{(\xi)}$ norm of $(f_i - \mathrm{proj}_{\{f_{i+1},\ldots,f_k\}} f_i)$, and since for each i there is (by linear independence) a ξ_i for which this squared norm is $> \epsilon > 0$, we clearly have

$$k_\lambda \geqslant \min_c K_\lambda\left(\sum_1^s \xi_i/s, c\right) > \epsilon/s > 0$$

for all λ. We hereafter write

$$K^{(i)}(x, c) = \left[f_i(x) - \sum_{j>i} c_{ij} f_j(x)\right]^2.$$

https://doi.org/10.4153/CJM-1962-050-8 Published online by Cambridge University Press

Let F be the following mapping from G into the $(s-1)$-dimensional simplex $E = \{(e_1, \ldots, e_s) : e_i \geqslant 0,\ 1 \leqslant i \leqslant s,\ \sum e_i = 1\}$:

$$F(\lambda, c_\lambda, \xi_\lambda) = \left(\frac{\lambda_1 K^{(1)}(\xi_\lambda, c_\lambda)}{k_\lambda}, \ldots, \frac{\lambda_s K^{(s)}(\xi_\lambda, c_\lambda)}{k_\lambda}\right).$$

Since $k_\lambda = \sum \lambda_j K^{(j)}(\xi_\lambda, c_\lambda)$ when ξ_λ is maximin and c_λ is minimax, the range of F is indeed in E. It is easy to see that F is continuous.

We want to show that F is onto. Obviously, for λ restricted to a vertex of Λ, F maps V_λ onto the corresponding vertex of E, and similarly F maps the part $\bigcup_{\lambda \in T} V_\lambda$ of G above any subsimplex (edge, face, etc.) T of Λ into the corresponding subsimplex of E. It follows from the theorem of Namioka **(3)** cited above that F is onto. (In general, $F(V_\lambda)$ need not be a point, which is why we could not directly define a mapping $F : \Lambda \to E$; there need not exist a cross-section from Λ into G, so we need Namioka's result.)

III. Since F is onto, there is a point $(\bar\lambda, c_{\bar\lambda}', \xi_{\bar\lambda}')$ of G which F maps into $(s^{-1}, s^{-1}, \ldots, s^{-1})$. Let ξ^* (of the statement of the theorem) $= \xi_{\bar\lambda}'$ and

$$g_i(x) = \left[f_i(x) - \sum_{j>i} c'_{\bar\lambda ij} f_j(x)\right] \Big/ [K^{(i)}(\xi'_{\bar\lambda}, c'_{\bar\lambda})]^{1/2}.$$

(The denominator, being $[k_{\bar\lambda}/\bar\lambda_i s]^{1/2}$ where $\bar\lambda \in \operatorname{Int}\Lambda$, is finite and not zero.) Clearly, $\int g_i^2 d\xi_{\bar\lambda}' = 1$; and since the $c_{\bar\lambda ij}'$ for each i are chosen to minimize $K^{(i)}(\xi_{\bar\lambda}', c)$, we see that the g_i are orthonormal $(\xi_{\bar\lambda}')$, $1 \leqslant i \leqslant s$, and are orthogonal $(\xi_{\bar\lambda}')$ to the f_j, $j > s$. Finally,

$$\sum_{i=1}^{s} g_i^2(x) = \sum_{i=1}^{s} \frac{K^{(i)}(x, c'_{\bar\lambda})}{K^{(i)}(\xi'_{\bar\lambda}, c'_{\bar\lambda})} = \sum_{i=1}^{s} \frac{s\bar\lambda_i K^{(i)}(x, c'_{\bar\lambda})}{k_{\bar\lambda}} = s\,\frac{K_{\bar\lambda}(x, c'_{\bar\lambda})}{K_{\bar\lambda}(\xi'_{\bar\lambda}, c'_{\bar\lambda})}.$$

By the game-theoretic results,

$$\max_x K_{\bar\lambda}(x, c'_{\bar\lambda}) = K_{\bar\lambda}(\xi'_{\bar\lambda}, c'_{\bar\lambda}),$$

proving the desired result.

3. Relationship to previous method and results. Let f denote the column vector of f_j's and $M(\xi)$ the matrix $\int f(x) f(x)' \xi(dx)$. It was shown in **(2)** that ξ^* maximizes $\det M(\xi)$ if and only if ξ^* minimizes $\max_x f(x)' M^{-1}(\xi) f(x)$ among ξ for which $M(\xi)$ is non-singular, and if and only if $\max_x f(x)' M^{-1}(\xi) f(x) = k$. This yields the theorem of the present paper when $s = k$. If $s < k$, let $\bar f$ be the last $k - s$ functions of f and $\bar M(\xi) = \int \bar f(x) \bar f(x)' \xi(dx)$. It was shown in **(1)** that, if $M(\xi^*)$ is non-singular, then ξ^* minimizes the determinant of the upper left-hand $s \times s$ submatrix of $M^{-1}(\xi)$ (which can be defined in an obvious way even if $M(\xi)$ is singular) if and only if ξ^* minimizes

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

and if and only if $d(\xi^*) = s$. When $s = k$, non-singularity is no restriction, but when $s < k$ the result of the present paper is now obtained in **(1)** only under the artificial *ad hoc* assumption that there exists a minimizing ξ^* for which $M(\xi^*)$ is non-singular. The result of **(1)** in the general case is obtained only by replacing $d(\xi)$ above by the quantity $\max_{\xi'} D(\xi', \xi)$, where

$$\begin{aligned} D(\xi', \xi) &= \bar{D}(\xi', \xi) - \rho(\xi', \xi) \\ &= \operatorname{tr}[\bar{J}^{-1}(\xi)\bar{J}(\xi')] - \operatorname{tr}[J_3^{-1}(\xi)J_3(\xi')] - \rho(\xi', \xi), \end{aligned}$$

where $J(\xi) = AM(\xi)\, A'$ is of rank $r + s$ with zero elements outside the first principal $(r + s) \times (r + s)$ submatrix $\bar{J}(\xi)$ and with A triangular (zeros below the main diagonal), $J_3(\xi)$ is the lower right-hand $r \times r$ submatrix of $\bar{J}(\xi)$, $\bar{J}(\xi')$ and $J_3(\xi')$ have the same meanings for the same A (for which $J(\xi')$ need not have the same properties as $J(\xi)$), and

$$\rho(\xi', \xi) = \operatorname{tr}[J_1^{-1}(\xi) \lim_{\lambda \to 0} J_4(\xi')[J_5(\xi') + \lambda I]^{-1} J_4(\xi')']$$

where J_1, J_5, and J_4 are, respectively, the first principal $s \times s$ submatrix of J, the last principal $(k - s - r) \times (k - s - r)$ submatrix of J, and the corresponding $s \times (k - s - r)$ submatrix of J.

Although D is invariantly defined (does not depend on the choice of A), $\bar{D}$ and ρ are not. Clearly, the condition $\max_{\xi'} \bar{D}(\xi', \xi) = s$ is sufficient for the result $\max_{\xi'} D(\xi', \xi) = s$, but it is not necessary (except under the *ad hoc* assumption, which implies that $\rho = 0$ and $\max_{\xi'} D(\xi', \xi^*) = d(\xi^*)$). The result of the present paper is precisely that, for ξ^* satisfying the above-mentioned extremum criteria of **(1)**, there is a choice of A for which $\max_{\xi'} \bar{D}(\xi', \xi^*) = s$. Conversely, this last result, combined with that of **(1)**, would yield the theorem of the present paper, but the methods of **(1)** do not yield this result on the choice of A. Thus, the present theorem states a somewhat stronger result than that of **(1)** in the general case. One can also try to obtain the present result by a passage to the limit from the case satisfying the extra assumption, but considerable delicacy is involved.

References

1. J. Kiefer, *Optimum designs in regression problems, II*, Ann. Math. Stat., *32* (1961), 298–325.

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

3. I. Namioka, *On certain onto maps*, Can. J. Math., *14* (1962), 461–466.

Cornell University