OPTIMAL EXPERIMENTAL DESIGNS[1]

By Samuel Karlin[2] and William J. Studden[3]

1. Introduction. The purpose of this paper is to discuss a number of results concerning the geometric theory of the optimal design of experiments. This theory was initiated and principally developed in a series of important publications by Elfving (1952), Kiefer (1960), (1962) and Kiefer and Wolfowitz (1959), (1960). For other references and historical background we direct the reader to Kiefer (1959), (1960). This paper was motivated and inspired by numerous conversations with Kiefer and contains approximately half expository and half new material. Almost all the proofs are new and presented in a unified manner.

The theory of the optimal design of experiments fits the following structure. Let $\mathbf{f} = (f_0, f_1, \cdots, f_n)$ denote a vector of $n + 1$ linearly independent continuous functions defined on a compact space $\mathfrak{X}$. The points of $\mathfrak{X}$ are referred to as possible levels of feasible experiments. For each level $x \varepsilon \mathfrak{X}$ some experiment can be performed whose outcome is a random variable $y(x)$. It is assumed that $y(x)$ has a mean of the explicit form[4] $\mathcal{E}y(x) = \sum_{j=0}^{n} \theta_j f_j(x)$ and a common variance σ^2 independent of x (normalized for convenience $= 1$). The functions $f_0, \cdots, f_n$, called the regression functions, are assumed known to the experimenter while the parameters $\theta_0, \theta_1, \cdots, \theta_n$ are unknowns to be estimated on the basis of N *uncorrelated* observations $\{y(x_i)\}_1^N$.

An *experimental design* specifies a probability measure ξ concentrating mass $p_1, \cdots, p_r$ at the points $x_1, \cdots, x_r$ where $p_i N = n_i$, $i = 1, \cdots, r$, are integers. The associated experiment involves taking n_i observations of the random variable $y(x_i)$, $i = 1, \cdots, r$.

The problem confronting the experimenter is to choose the design possessing certain optimality properties. Statistical considerations [see Kiefer (1959)] direct an interest in the matrix $\mathbf{M}(\xi) = \|m_{ij}(\xi)\|_{i,j=0}^{n}$ $(m_{ij}(\xi) = \int_{\mathfrak{X}} f_i(x) f_j(x) \xi(dx))$ commonly called the *information matrix* of the design ξ. If the unknown parameter vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_n)$ is estimated by the method of least squares thus securing a best linear unbiased estimate, say $\hat{\boldsymbol{\theta}}$, then the covariance matrix of $\hat{\boldsymbol{\theta}}$ is given by

$$\mathcal{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' = N^{-1}\mathbf{M}^{-1}(\xi) \tag{1.1}$$

where ξ assigns mass $p_i = n_i/N$ at the points x_i, $i = 1, \cdots, r$. If the matrix $\mathbf{M}^{-1}(\xi)$ is "small" or $\mathbf{M}(\xi)$ is "large", then roughly speaking $\hat{\boldsymbol{\theta}}$ is close to $\boldsymbol{\theta}$. Most

Received 29 September 1965.

[1] Reproduction in whole or in part permitted for any purposes of the United States Government.

[2] Work supported in part by Contracts NONR 225(28) and NSF GP 2487 at Stanford University.

[3] Work supported in part by Contract NONR 1100(26) at Purdue University.

[4] $\mathcal{E}$ denotes the expectation operator.

criteria for discerning optimality of an experimental design are based on maximizing some appropriate functional of the matrix $\mathbf{M}(\xi)$.

In the following we shall assume that ξ is an *arbitrary* probability measure on the Borel sets $\mathcal{B}$ of $\mathcal{X}$ where $\mathcal{B}$ includes all one point sets. The justification and convenience in allowing this greater generality in the choice of measures ξ is that it permits a complete characterization of certain optimal designs. Of course, any choice of a probability measure can usually be implemented in practice by closely related bona fide designs.

A simple measure of the magnitude of the information matrix $\mathbf{M}(\xi)$ is its determinant $|\mathbf{M}(\xi)|$. A design ξ^* is said to be D-optimal if ξ^* maximizes $|\mathbf{M}(\xi)|$. Another criterion for optimality, formalized and interpreted in Kiefer (1959), is as follows. Let

$$d(x, \xi) = \sup_{\mathbf{d}} [(\mathbf{f}(x), \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})] \tag{1.2}$$

where the ratio on the right is defined to be zero whenever the denominator vanishes. (The notation $(\mathbf{u}, \mathbf{v})$ signifies the inner product of the vectors $\mathbf{u}$ and $\mathbf{v}$.) In case $|\mathbf{M}(\xi)| \neq 0$ the function $d(x, \xi)$ reduces to the more familiar expression $d(x, \xi) = (\mathbf{f}(x), \mathbf{M}^{-1}(\xi)\mathbf{f}(x))$. The expression $N^{-1} d(x, \xi)$ is the variance of the best linear unbiased estimate of the regression function $(\mathbf{f}(x), \boldsymbol{\theta}) = \sum_{i=0}^{n} \theta_i f_i(x)$.

The following result is due to Kiefer and Wolfowitz (1960), (see Section 4).

THEOREM 4.1. (Equivalence Theorem). *The three classes of probability measures ξ^* defined by*

(i) ξ^* *maximizes* $|\mathbf{M}(\xi)|$,

(ii) ξ^* *minimizes* $\sup_x d(x, \xi)$,

(iii) $\sup_x d(x, \xi^*) = n + 1$

coincide. The set Γ of all ξ^* satisfying these conditions is convex and closed and $\mathbf{M}(\xi^*)$ is the same for all $\xi^* \in \Gamma$.

The functions $|\mathbf{M}(\xi)|$ and $d(x, \xi)$ are relevant for the problem of estimating the full set of $n + 1$ parameters $\theta_0, \theta_1, \cdots, \theta_n$. Kiefer (1962) has partially extended the equivalence theorem in a rather complicated manner to the case of estimating a subset of the parameters.

In considering the estimation of $s + 1$ $(s < n)$ of the parameters $\theta_0, \theta_1, \cdots, \theta_n$ we introduce two functions serving as the counterparts of the functions $|\mathbf{M}(\xi)|$ and $d(x, \xi)$. When $\mathbf{M}(\xi)$ is nonsingular, an analog of $|\mathbf{M}(\xi)|$ is traditionally taken to be

$$|\mathbf{M}_s^*(\xi)| = |\mathbf{M}_1(\xi) - \mathbf{M}_2'(\xi)\mathbf{M}_3^{-1}(\xi)\mathbf{M}_2(\xi)|$$

where

$$\mathbf{M}(\xi) = \begin{Vmatrix} \mathbf{M}_1(\xi) & \mathbf{M}_2'(\xi) \\ \mathbf{M}_2(\xi) & \mathbf{M}_3(\xi) \end{Vmatrix}, \qquad (\mathbf{M}_1(\xi) \text{ is } s + 1 \times s + 1)$$

and the corresponding version of $d(x, \xi)$ is defined to be

$$d_s(x, \xi) = ((\mathbf{f}^{(1)}(x) - \mathbf{D}'(\xi)\mathbf{f}^{(2)}(x)), \mathbf{P}(\xi)(\mathbf{f}^{(1)}(x) - \mathbf{D}'(\xi)\mathbf{f}^{(2)}(x)))$$

where $\mathbf{f}^{(1)}(x) = (f_0(x), \cdots, f_s(x))$, $\mathbf{f}^{(2)}(x) = (f_{s+1}(x), \cdots, f_n(x))$, $\mathbf{P}(\xi) = [M_s^*(\xi)]^{-1}$ and $\mathbf{D}(\xi) = \mathbf{M}_3^{-1}(\xi)\mathbf{M}_2(\xi)$.

The above definitions will be properly extended to the case where $\mathbf{M}_3(\xi)$ is singular (see Section 6). One of the principal theorems of this paper is the following game theoretic result which is basic in establishing the equivalence of various criteria for estimating $s + 1$ out of the $n + 1$ parameters.

THEOREM 6.1. *Let $\mathcal{P}$ denote the set of all positive definite matrices* $\mathbf{P}$ (*of order* $(s + 1) \times (s + 1)$) *normalized so that* $|\mathbf{P}|^{-1} = \sup_\xi |\mathbf{M}_s(\xi)|$ *and let $\mathfrak{D}$ comprise the set of all real matrices* $\mathbf{D}$ *of order* $(n - s) \times (s + 1)$. *If*

$$\phi(\mathbf{P}, \mathbf{D}; \xi) = \int (\mathbf{f}^{(1)}(x) - \mathbf{D}'\mathbf{f}^{(2)}(x), \mathbf{P}(\mathbf{f}^{(1)}(x) - \mathbf{D}'\mathbf{f}^{(2)}(x))\xi(dx)$$

then

$$\sup_\xi \inf_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi) = s + 1 = \inf_{\mathbf{P},\mathbf{D}} \sup_\xi \phi(\mathbf{P}, \mathbf{D}; \xi).$$

Furthermore, there exists a probability measure ξ_0 satisfying $|\mathbf{M}_s(\xi_0)| = \sup_\xi |\mathbf{M}_s(\xi)|$ *and*

$$\phi(\mathbf{P}, \mathbf{D}; \xi_0) \geqq s + 1 \quad \text{for all} \quad (\mathbf{P}, \mathbf{D}) \in \mathcal{P} \times \mathfrak{D}.$$

If $(\mathbf{P}_0, \mathbf{D}_0) \in \mathcal{P} \times \mathfrak{D}$ *fulfills*

$$\phi(\mathbf{P}_0, \mathbf{D}_0; \xi) \leqq s + 1 \quad \text{for all probability measures} \quad \xi \quad \text{on} \quad \mathfrak{X}$$

then $\mathbf{D}_0$ *satisfies* $\mathbf{M}_3(\xi_0)\mathbf{D}_0 = \mathbf{M}_2(\xi_0)$ *and*

$$\mathbf{P}_0 = [\mathbf{M}_1(\xi_0) - \mathbf{D}_0'\mathbf{M}_3(\xi_0)\mathbf{D}_0]^{-1}.$$

As a consequence of the above theorem we deduce [compare with Kiefer (1962)]:

THEOREM 6.2. *The following three conditions are equivalent;*

(i) ξ^* *maximizes* $|\mathbf{M}_s^*(\xi)|$,
(ii) ξ^* *minimizes* $\sup_x d_s(x; \xi)$,
(iii) $\sup_x d_s(x; \xi^*) = s + 1$.

In the problem of estimating all of the parameters $\theta_0, \theta_1, \cdots, \theta_n$ we shall exploit a theorem due to Schoenberg which states that when $f_i(x) = [w(x)]^{\frac{1}{2}}x^i$, $i = 0, 1, \cdots, n$, and $w(x)$ is a classical type weight function then $|\mathbf{M}(\xi)|$ is maximized provided ξ concentrates equal mass at $n + 1$ points which are the zeros of certain familiar orthogonal polynomials. As an example, we cite

THEOREM 5.1. *Let* $f_i(x) = [w(x)]^{\frac{1}{2}}x^i$, $i = 0, 1, \cdots, n$, *where* $w(x) = (1 - x)^{\alpha+1}(1 + x)^{\beta+1}$ *and* $\mathfrak{X} = [-1, 1]$. *Then* $\max_\xi |\mathbf{M}(\xi)|$ *is uniquely achieved by the measure concentrating equal mass at the zeros of the* $n + 1$st *Jacobi polynomial* $P_n^{(\alpha,\beta)}(x)$.

A brief section by section outline of the paper is as follows: In Section 2 we consider the estimation of linear forms $(\mathbf{c}, \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_n)$ is the vector of unknown regression coefficients. Theorem 2.1 provides an exact expression of the minimum variance among all linear unbiased estimates of $(\mathbf{c}, \boldsymbol{\theta})$ based on a fixed design ξ, even allowing the contingency that $\mathbf{M}(\xi)$ may be singular. Theorem

2.2 embraces Elfving's geometric characterization of the optimal design for estimating $(\mathbf{c}, \boldsymbol{\theta})$. The theorem of Elfving has been slightly extended utilizing game theoretic methods. We shall include complete proofs of these two theorems since the majority of discussions of these results are loosely written where rank conditions on the relevant information matrices are usually ignored.

In Section 3 we consider a recent result of Hoel and Levine (1964) on extrapolation. We will show that the Hoel and Levine result and some of its extensions considered by Kiefer and Wolfwitz (1965) can be deduced directly from the Elfving result of Theorem 2.2.

Section 4, 5 and 6 investigate various minimax designs. Section 4 is devoted to a game theoretic proof of the equivalence theorem while in Section 5 we determine several explicit minimax designs for certain cases where the regression functions are related to classical orthogonal polynomials. Extensions of the results of Section 4 to the case of the estimation of a subset of the regression parameters $\theta_0, \cdots, \theta_n$ are given in Section 6.

Section 7 is devoted to a discussion of some classes of admissible designs. Some refinements and extensions of results in Kiefer (1959) are indicated. In the final Section 8 we consider some quadratic problems involving the minimization of $\mathcal{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'$ where $\mathbf{L}$ is a general nonnegative (i.e., positive semi-definite) matrix.

2. Estimation of the linear form $(\mathbf{c}, \boldsymbol{\theta})$. In this section we shall present two theorems concerning unbiased linear estimation of a linear form $(\mathbf{c}, \boldsymbol{\theta}) = \sum_{i=0}^{n} c_i\theta_i$. The first theorem exhibits an expression for the minimum variance among all linear unbiased estimates of $(\mathbf{c}, \boldsymbol{\theta})$ when a fixed design ξ is used. The second theorem is Elfving's characterization of the optimal (see Definition 2.1) design ξ for estimating $(\mathbf{c}, \boldsymbol{\theta})$.

Let $\mathbf{f}(x) = (f_0(x), f_1(x), \cdots, f_n(x))$ be a vector-valued function composed of $n + 1$ linearly independent continuous functions defined on a compact space $\mathfrak{X}$. Let ξ denote an arbitrary probability measure defined on the Borel field $\mathfrak{B}$ generated by the open sets of $\mathfrak{X}$. For each ξ let $\mathbf{M}(\xi)$ denote the matrix

$$\mathbf{M}(\xi) = \|m_{ij}(\xi)\|_{i,j=0}^{n} \tag{2.1}$$

where

$$m_{ij}(\xi) = \int_{\mathfrak{X}} f_i(x)f_j(x)\xi(dx), \qquad i, j = 0, 1, \cdots, n. \tag{2.2}$$

Throughout we assume that $\mathfrak{B}$ includes all one point sets. In the case where the measure ξ_0 concentrates at a single point x_0 the matrix $\mathbf{M}(\xi_0)$ reduces to a rank one matrix with elements $m_{ij} = f_i(x_0)f_j(x_0)$. We also single out the case where ξ concentrates masses $p_1, \cdots, p_r$ at the r distinct points $x_1, \cdots, x_r$ respectively. Here $\mathbf{M}(\xi)$ takes the form

$$\sum_{l=1}^{r} p_l[\mathbf{f}(x_l)][\mathbf{f}(x_l)]' \tag{2.3}$$

whose individual elements are $m_{ij} = \sum_{l=1}^{r} p_l f_i(x_l)f_j(x_l)$.

In the case where $f_i(x) = x^i$, $i = 0, \cdots, n$, and $\mathfrak{X} = [a, b]$ the matrices $\mathbf{M}(\xi)$ reduce to the classical Hankel matrices $\|c_{i+j}\|_{i,j=0}^{n}$ where $c_k = \int_a^b x^k \xi(dx)$, $k = 0, 1, \cdots, 2n$.

In the following lemma we record several familiar properties of the matrices $\mathbf{M}(\xi)$.

LEMMA 2.1. *Let* $\mathbf{M}(\xi)$ *be defined as in* (2.1) *and* (2.2). *Then*

(i) *for each* ξ, $\mathbf{M}(\xi)$ *is positive semi-definite;*

(ii) $|\mathbf{M}(\xi)| = 0$ *whenever the spectrum of* ξ *contains less than* $n + 1$ *points;*

(iii) *the family of matrices* $\mathbf{M}(\xi)$, *as* ξ *ranges over the class of probability measures, is a convex compact set;*

(iv) *for each* ξ *the matrix* $\mathbf{M}(\xi)$ *may be written in the form* (2.3) *where* $r \leqq (n + 1)(n + 2)/2 + 1$.

PROOF. (i) For any choice $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)$ with α_i all real we have

$$(\boldsymbol{\alpha}, \mathbf{M}(\xi)\boldsymbol{\alpha}) = \sum_{i,j=0}^{n} m_{ij}(\xi)\alpha_i\alpha_j = \int_{\mathfrak{X}} \left|\sum_{i=0}^{n} \alpha_i f_i(x)\right|^2 \xi(dx) \geqq 0$$

which shows that $\mathbf{M}(\xi)$ is positive semi-definite.

(ii) Under the hypothesis stated the rank of $\mathbf{M}(\xi)$ is at most n and therefore $\mathbf{M}(\xi)$ is singular.

(iii) The convexity assertion is immediate.

Let $z_1(x), \cdots, z_m(x)$, $(m = (n + 1)(n + 2)/2)$ denote the set of functions $f_i(x)f_j(x)$ $(j = 0, 1, \cdots, i;\ i = 0, 1, \cdots, n)$ arranged in some order and consider the set C_m in E^m generated by the coordinate functions, i.e., $C_m = \{(z_1(x), \cdots, z_m(x)) \mid x \in \mathfrak{X}\}$. Let $\mathcal{C}(C_m)$ denote the convex hull of C_m. A classical theorem of Caratheodory,[5] informs us that each $\mathbf{c} = (c_1, \cdots, c_m)$ in $\mathcal{C}(C_m)$ possesses a representation of the form

$$c_i = \sum_{j=1}^{m+1} \alpha_j z_i(t_j), \qquad i = 1, 2, \cdots, m, \tag{2.4}$$

where $\alpha_j \geqq 0$, $j = 1, \cdots, m + 1$ and $\sum_{j=1}^{m+1} \alpha_j = 1$. Since the functions $z_1, \cdots, z_m$ are continuous and $\mathfrak{X}$ is compact we deduce from (2.4) that $\mathcal{C}(C_m)$ is compact. Furthermore, the fact that $\mathcal{C}(C_m)$ is closed implies $\mathcal{C}(C_m) = \mathfrak{M}_m$ where

$$\mathfrak{M}_m = \{(c_1, \cdots, c_m) \mid c_i = \int z_i(x)\xi(dx), \qquad i = 1, \cdots, m\}.$$

The compactness of the family of matrices $\mathbf{M}(\xi)$ is now clear.

(iv) This is nothing more than an expression of the Caratheodory theorem for the case at hand.

We now direct attention to the problem of estimating a linear form $(\mathbf{c}, \boldsymbol{\theta})$. Let $\xi \equiv \{x_i; p_i\}_1^r$ where $p_iN = n_i$ are integers; i.e., the design ξ involves taking n_i observations at the level x_i, $i = 1, \cdots, r$. In considering unbiased estimates, the linear form $(\mathbf{c}, \boldsymbol{\theta})$ must of course be estimable when the design ξ is used, i.e., there must exist on N-vector $\boldsymbol{\gamma}$ such that $\mathcal{E}(\boldsymbol{\gamma}, \mathbf{y}) = (\mathbf{c}, \boldsymbol{\theta})$ for all $\boldsymbol{\theta}$, where $\mathbf{y}$ denotes the N-vector of observations $\{y(x_i)\}_1^N$. Let $\mathbf{A}$ denote a matrix of order $N \times (n + 1)$

[5] The theorem of Caratheodory in question states that every point in the smallest convex set containing a given set A in Euclidean m space can be represented as a convex combination of at most $m + 1$ points of A.

with $n_i = Np_i$ of the rows equal to $\mathbf{f}(x_i) = (f_0(x_i), \cdots, f_n(x_i)), i = 1, \cdots, r$. Note that $\mathbf{A}'\mathbf{A} = N\mathbf{M}(\xi)$. If $(\boldsymbol{\gamma}, \mathbf{y})$ is an unbiased estimate of $(\mathbf{c}, \boldsymbol{\theta})$ then $(\mathbf{c}, \boldsymbol{\theta}) = \mathcal{E}(\boldsymbol{\gamma}, \mathbf{y}) = (\boldsymbol{\gamma}, \mathbf{A}\boldsymbol{\theta}) = (\mathbf{A}'\boldsymbol{\gamma}, \boldsymbol{\theta})$ for all $\boldsymbol{\theta}$ so that $\mathbf{A}'\boldsymbol{\gamma} = \mathbf{c}$. Thus for a given design $\xi = \{x_i ; p_i\}_1^r$ the vector $\mathbf{c}$ is estimable only if there exists a solution $\boldsymbol{\gamma}$ of the equation $\mathbf{A}'\boldsymbol{\gamma} = \mathbf{c}$, i.e., $\mathbf{c}$ belongs to the range of $\mathbf{A}'$.

The following theorem provides a convenient expression for the minimum variance of any linear unbiased estimate of $(\mathbf{c}, \boldsymbol{\theta})$ obtained from a fixed design ξ.

THEOREM 2.1. *Let $\xi = \{x_i ; p_i\}_1^r$ where $p_iN = n_i$ are integers and let $\mathfrak{F}(\xi)$ denote the class of linear unbiased estimates $(\boldsymbol{\gamma}, \mathbf{y})$ of $(\mathbf{c}, \boldsymbol{\theta})$. If $\mathfrak{F}(\xi)$ is non-void,* i.e., $\mathbf{c}$ *is estimable, and $V((\boldsymbol{\gamma}, \mathbf{y}))$ denotes the variance of $(\boldsymbol{\gamma}, \mathbf{y})$ then*

$$V((\boldsymbol{\gamma}_0, \mathbf{y})) = \min_{\mathfrak{F}(\xi)} V((\boldsymbol{\gamma}, \mathbf{y})) = N^{-1} \sup_{\mathbf{d} \varepsilon U^{\perp}, \mathbf{d} \neq 0} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})]$$
$$= N^{-1} \textstyle\sum_1^s [(\boldsymbol{\varphi}_i, \mathbf{c})^2/\lambda_i]$$

*where $U = \{\mathbf{d} \mid \mathbf{M}(\xi)\mathbf{d} = \mathbf{0}\}$, $\boldsymbol{\gamma}_0 = N^{-1}\mathbf{A}\mathbf{d}_0$, $\mathbf{d}_0 = \sum_{i=1}^s \lambda_i^{-1}(\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}_i$ and $\lambda_1, \cdots, \lambda_s$ are the non-zero eigenvalues of $\mathbf{M}(\xi)$ with associated orthonormal eigen-*vectors $\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_s$.

PROOF. The variance $V((\boldsymbol{\gamma}, \mathbf{y}))$ of any estimate $(\boldsymbol{\gamma}, \mathbf{y}) \varepsilon \mathfrak{F}(\xi)$ is easily seen to be $(\boldsymbol{\gamma}, \boldsymbol{\gamma})$. Moreover

$$(2.5) \qquad (\mathbf{c}, \mathbf{d})^2 = (\mathbf{A}'\boldsymbol{\gamma}, \mathbf{d})^2 = (\boldsymbol{\gamma}, \mathbf{A}\mathbf{d})^2 \leqq N(\boldsymbol{\gamma}, \boldsymbol{\gamma})(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})$$

so that

$$(2.6) \qquad V((\boldsymbol{\gamma}, \mathbf{y})) \geqq N^{-1} \sup_{\mathbf{d} \varepsilon U^{\perp}, \mathbf{d} \neq 0} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})].$$

From (2.5) it follows that $\mathbf{c} \varepsilon U^{\perp}$ and hence $\mathbf{c} = \sum_1^s (\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}$. Therefore for any $\mathbf{d} \varepsilon U^{\perp}$ we have

$$\begin{aligned}(\mathbf{c}, \mathbf{d})^2 &= (\textstyle\sum_{i=1}^s (\boldsymbol{\varphi}_i, \mathbf{d})(\boldsymbol{\varphi}_i, \mathbf{c}))^2 \\ (2.7) \qquad &= [\textstyle\sum_{i=1}^s (\lambda_i)^{\frac{1}{2}}(\boldsymbol{\varphi}_i, \mathbf{d})(\boldsymbol{\varphi}_i, \mathbf{c})/(\lambda_i)^{\frac{1}{2}}]^2 \\ &\leqq \textstyle\sum_{i=1}^s \lambda_i(\boldsymbol{\varphi}_i, \mathbf{d})^2 \sum_{i=1}^s (\boldsymbol{\varphi}_i, \mathbf{c})^2/\lambda_i \\ &= (\mathbf{d}, \mathbf{M}(\xi)\mathbf{d}) \textstyle\sum_1^s (\boldsymbol{\varphi}_i, \mathbf{c})^2/\lambda_i \qquad (\mathbf{M}(\xi) = \sum_{i=1}^s \lambda_i\boldsymbol{\varphi}_i\boldsymbol{\varphi}_i').)\end{aligned}$$

Equality occurs if and only if $\lambda_i(\boldsymbol{\varphi}_i, \mathbf{d}) = k(\boldsymbol{\varphi}_i, \mathbf{c})$ $(i = 1, 2, \cdots, s)$ for some constant k or when $\mathbf{d}$ is proportional to $\mathbf{d}_0 = \sum_{i=1}^s \lambda_i^{-1}(\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}_i$. Now if $\boldsymbol{\gamma}_0 = N^{-1}\mathbf{A}\mathbf{d}_0$ then equality takes place in (2.5) and

$$(2.8) \qquad \mathbf{A}'\boldsymbol{\gamma}_0 = \mathbf{M}(\xi)\mathbf{d}_0 = \textstyle\sum_{i=1}^s (\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}_i = \mathbf{c}.$$

Thus the bound in (2.6) is attainable by the estimate $(\boldsymbol{\gamma}_0, \mathbf{y})$ which belongs to $\mathfrak{F}(\xi)$. This completes the proof of the theorem.

REMARK 2.1. When $\mathbf{M}(\xi)$ is non-singular we have available the more familiar expression $\sup_{\mathbf{d} \neq 0} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})] = (\mathbf{c}, \mathbf{M}^{-1}(\xi)\mathbf{c})$ achieved for $\mathbf{d}_0 = \mathbf{M}^{-1}(\xi)\mathbf{c}$. The matrix $\mathbf{A}$ $(\mathbf{A}'\mathbf{A} = N\mathbf{M}(\xi))$ in this case is necessarily of rank $n + 1$ and every vector $\mathbf{c}$ is estimable.

Suppose now that ξ is an arbitrary probability measure on $\mathfrak{X}$. A vector $\mathbf{c}$ is said to be *estimable with respect to the design* ξ if and only if $\mathbf{c}$ belongs to the linear space spanned by the set of vectors $\{\mathbf{f}(x) | x \varepsilon S(\xi) = \text{spectrum of } \xi\}$. Note that this definition is consistent with the usual definition when $\xi = \{x_i ; p_i\}_1^r$ and $p_i N = n_i$, $i = 1, \cdots, r$, are integers. For an arbitrary design ξ we define

$$d(\mathbf{c}, \xi) = \sup_{\mathbf{d} \varepsilon \mathbf{U}^\perp, \mathbf{d} \neq 0} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})]$$

whenever $\mathbf{c}$ is estimable with respect to ξ and ∞ otherwise. If $\mathbf{c} = \mathbf{f}(x)$ for some fixed value of $x \varepsilon \mathfrak{X}$ we shall write $d(x, \xi)$ for $d(\mathbf{c}, \xi)$.

DEFINITION 2.1. A design ξ is said to be optimal with respect to the estimation of $(\mathbf{c}, \boldsymbol{\theta})$ if ξ minimizes $d(\mathbf{c}, \xi)$.

The following result due to Elfving (1952) characterizes optimal designs ξ for the problem of estimating $(\mathbf{c}, \boldsymbol{\theta})$. This result is geometric in nature and for the cases $n = 1$ or 2 provides a simple explicit means of determining the optimal design.

THEOREM 2.2. *Let* $\mathcal{R}_+ = \{\mathbf{f}(x) = (f_0(x), \cdots, f_n(x)) \mid x \varepsilon \mathfrak{X}\}$ *and let* $\mathcal{R}_-$ *denote the symmetric image of* $\mathcal{R}_+$, i.e., $\mathcal{R}_- = \{-\mathbf{f}(x) \mid x \varepsilon \mathfrak{X}\}$. *Further, let* $\mathcal{R}$ *denote the convex hull of* $\mathcal{R}_+ \cup \mathcal{R}_-$. *A design* ξ_0 *is optimal with respect to the estimation of* $(\mathbf{c}, \boldsymbol{\theta})$ *if and only if there exists a real measurable function* $\varphi(x)$ *satisfying* $|\varphi(x)| \equiv 1$ *such that* $\mathbf{c}^* = \int \varphi(x)\mathbf{f}(x)\xi_0(dx)$ *is* (i) *proportional to* $\mathbf{c}$ *and* (ii) *a boundary point of* $\mathcal{R}$. *Moreover* $\beta\mathbf{c}$ *lies on the boundary of* $\mathcal{R}$ *if and only if* $\beta^2 = v_0^{-1}$ *where* $v_0 = \min_\xi d(\mathbf{c}, \xi)$.

PROOF. We first assume that ξ_0 is optimal and show that conditions (i) and (ii) are satisfied. Consider the sequence of games with kernel

$$\phi_\epsilon(\mu, \xi) = \int [(\mathbf{c}, \mathbf{g})^2/(\mathbf{g}, \mathbf{M}(\xi)\mathbf{g})]\mu(d\mathbf{g})$$

where Player I maximizes over the set of probability measures μ defined on the unit sphere S^{n+1} in E^{n+1} and Player II minimizes over the convex set of probability measures ξ on $\mathfrak{X}$ such that $\mathbf{M}(\xi)$ has all of its eigenvalues at least ϵ. It is easy to verify that the kernel ϕ is convex in ξ and linear in μ. Since the strategy spaces are compact each game has a value v_ϵ and associated optimal stategies $\mathbf{M}(\xi_\epsilon)$ and μ_ϵ. Furthermore

$$v_\epsilon = \sup_{\mathbf{d} \varepsilon S^{n+1}} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi_\epsilon)\mathbf{d})]$$

is uniquely achieved for $\mathbf{d}$ proportional to $\mathbf{M}^{-1}(\xi_\epsilon)\mathbf{c}$. It follows that μ_ϵ concentrates on a single point $\mathbf{d}_\epsilon$. A straightforward compactness argument shows that there exists a $\mathbf{d}^*$ such that

$$(\mathbf{c}, \mathbf{d}^*)^2 \geqq v_0(\mathbf{d}^*, \mathbf{M}(\xi)\mathbf{d}^*) \qquad \text{for all} \qquad \xi$$

where $v_0 = \inf_\xi \sup_{\mathbf{d} \varepsilon S^{n+1}} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})]$.

Letting ξ concentrate on a single point x we infer that

$$(2.9) \qquad (\mathbf{c}, \mathbf{d}^*)^2 \geqq v_0(\mathbf{d}^*, \mathbf{f}(x))^2 \qquad \text{for all} \quad x \varepsilon \mathfrak{X}.$$

For $\xi = \xi_0$ we also have

$$(2.10) \qquad (\mathbf{c}, \mathbf{d}^*)^2 \geqq v_0(\mathbf{d}^*, \mathbf{M}(\xi_0)\mathbf{d}^*).$$

But ξ_0 is optimal with respect to the estimation of $(\mathbf{c}, \boldsymbol{\theta})$ so that

(2.11) $v_0 = \inf_\xi \sup_{\mathbf{d} \varepsilon S^{n+1}} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})] = \sup_{\mathbf{d} \varepsilon S^{n+1}} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi_0)\mathbf{d})]$.

Comparing (2.10) and (2.11) we see that equality holds in (2.10) and hence in (2.9) for $x \varepsilon S(\xi_0)$ = spectrum of ξ_0. Thus

$$(\mathbf{c}, \mathbf{d}^*)^2 = v_0(\mathbf{d}^*, \mathbf{f}(x))^2, \qquad x \varepsilon S(\xi_0). \tag{2.12}$$

We now take a real square root of (2.12) to obtain

$$\varphi(x)(\mathbf{c}, \mathbf{d}^*) = (v_0)^{\frac{1}{2}}(\mathbf{d}^*, \mathbf{f}(x)), \qquad x \varepsilon S(\xi_0),$$

where $|\varphi(x)| \equiv 1$. We define $\varphi(x) = 1$ for $x \not\varepsilon S(\xi_0)$. Since each $f_i(x)$ $(i = 0, 1, \cdots, n)$ is continuous the function $\varphi(x)$ is clearly measurable.

Now from (2.7) we know that equality occurs in (2.10) only if the component of $\mathbf{d}^*$ in $U^\perp$ $(U = \{\mathbf{d} \mid \mathbf{M}(\xi_0)\mathbf{d} = \mathbf{0}\})$ is proportional to $\sum_{i=1}^s \lambda_i^{-1}(\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}_i$ where $\lambda_1, \cdots, \lambda_s$ are the non-zero eigenvalues of $\mathbf{M}(\xi_0)$ with corresponding orthonormal eigenvectors $\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_s$. By adjusting the proportionality constant we may assume $\mathbf{d}^*$ satisfies $\mathbf{M}(\xi_0)\mathbf{d}^* = \mathbf{M}(\xi_0)(\sum_{i=1}^s \lambda_i^{-1}(\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}_i) = \sum_{i=1}^s (\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}_i = \mathbf{c}$. Then $(\mathbf{c}, \mathbf{d}^*)^2 = v_0(\mathbf{d}^*, \mathbf{M}(\xi_0)\mathbf{d}^*) = v_0(\mathbf{d}^*, \mathbf{c})$. But $(\mathbf{c}, \mathbf{d}^*) = (\sum_{i=1}^s (\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}_i, \sum_{i=1}^s \lambda_i^{-1}(\boldsymbol{\varphi}_i, \mathbf{c})\boldsymbol{\varphi}_i) = \sum_{i=1}^s \lambda_i^{-1}(\boldsymbol{\varphi}_i, \mathbf{c})^2 > 0$ so that we may conclude that $(\mathbf{c}, \mathbf{d}^*) = v_0$. Therefore

$$\begin{aligned} \mathbf{c} = \mathbf{M}(\xi_0)\mathbf{d}^* &= \int (\mathbf{f}(x), \mathbf{d}^*)\mathbf{f}(x)\xi_0(dx) \\ &= v_0^{-\frac{1}{2}}(\mathbf{c}, \mathbf{d}^*) \int \varphi(x)\mathbf{f}(x)\xi_0(dx) \\ &= v_0^{\frac{1}{2}} \int \varphi(x)\mathbf{f}(x)\xi_0(dx). \end{aligned}$$

We have thus shown that $\int \varphi(x)\mathbf{f}(x)\xi_0(dx) = \beta\mathbf{c}$ where $\beta = v_0^{-\frac{1}{2}}$ which establishes condition (i) of the theorem.

We now verify that the vector $\beta\mathbf{c}$ $(\beta = v_0^{-\frac{1}{2}})$ lies on the boundary of $\mathcal{R}$. Since $\beta\mathbf{c} = \int \varphi(x)\mathbf{f}(x)\xi_0(dx)$ it follows that $\beta\mathbf{c} \varepsilon \mathcal{R}$. If $\beta\mathbf{c}$ is not on the boundary of $\mathcal{R}$ we choose $\sum_{\nu=1}^r \epsilon_\nu q_\nu \mathbf{f}(s_\nu) = \beta_1\mathbf{c}$ $(\epsilon_\nu = \pm 1, q_\nu > 0, \sum q_\nu = 1)$ on the boundary where $\beta_1 > \beta$. By Schwartz's inequality

$$\begin{aligned} \sup_{\mathbf{d} \varepsilon S^{n+1}} [(\mathbf{c}, \mathbf{d})^2/\textstyle\sum_{\nu=1}^r q_\nu(\mathbf{f}(s_\nu), \mathbf{d})^2] &\leqq \sup_{\mathbf{d} \varepsilon S^{n+1}} [(\mathbf{c}, \mathbf{d})^2/(\textstyle\sum_{\nu=1}^r \epsilon_\nu q_\nu(\mathbf{f}(s_\nu), \mathbf{d}))^2 \\ &= \beta_1^{-2} < \beta^{-2} = v_0 \end{aligned}$$

which contradicts (2.11). Therefore condition (ii) holds.

Since the set $\mathcal{R}$ is symmetric about the origin the above reasoning establishes that $\beta\mathbf{c}$ is on the boundary of $\mathcal{R}$ if and only if $\beta^2 = v_0^{-1}$.

To prove the converse half of the theorem we postulate the representation $\int \varphi(x)\mathbf{f}(x)\xi_0(dx) = \beta\mathbf{c}$ and that $\beta\mathbf{c}$ lies on the boundary of $\mathcal{R}$. Clearly $\mathbf{c}$ is estimable with respect to ξ_0 and $\beta^2 = v_0^{-1}$. Moreover, by Schwartz's inequality

$$\begin{aligned} (\mathbf{c}, \mathbf{d})^2 &= \beta^{-2}[\textstyle\int \varphi(x)(\mathbf{f}(x), \mathbf{d})\xi_0(dx)]^2 \\ &\leqq \beta^{-2} \textstyle\int (\mathbf{f}(x), \mathbf{d})^2\xi_0(dx) \\ &= \beta^{-2}(\mathbf{d}, \mathbf{M}(\xi_0)\mathbf{d}) \end{aligned}$$

and hence

$$d(\mathbf{c}, \xi_0) = \sup_{\mathbf{d}} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi_0)\mathbf{d})] \leqq \beta^{-2} = v_0 . \tag{2.13}$$

Since $v_0 = \min_\xi d(\mathbf{c}, \xi)$, equality must hold in (2.13) and hence ξ_0 is an optimal design with respect to the estimation of $(\mathbf{c}, \boldsymbol{\theta})$.

3. Some aspects of extrapolation. In this section we consider a theorem of Kiefer and Wolfowitz (1965) which generalizes a result due to Hoel and Levine (1964) concerning extrapolation in polynomial regression. It will be shown that the Hoel and Levine result and parts of Kiefer and Wolfowitz are direct consequences of Elfving's theorem. In the simple case when $f_i(x) = x^i$, $i = 0, 1, \cdots, n$, and $\mathfrak{X} = [-1, 1]$ the problem is to characterize the design minimizing the variance of the best linear unbiased estimate of the regression function $\sum_{i=0}^n \theta_i x_0^i$ where $|x_0| > 1$. This is called an extrapolation problem since the point x_0 is exterior to the interval $[-1, 1]$ on which the levels of the designs ξ are concentrated. In our previous terminology we wish to determine the optimal design with respect to the estimation of the linear form $(\mathbf{c}, \boldsymbol{\theta})$ for $\mathbf{c} = (1, x_0, x_0^2, \cdots, x_0^n)$.

The Hoel and Levine result asserts that the optimum design for estimating the linear form $(\boldsymbol{\theta}, \mathbf{f}(x_0)) = \sum_{i=0}^n \theta_i x_0^i$ concentrates on a *fixed* set of $n + 1$ points $-1 = s_0 < s_1 < \cdots < s_n = 1$ independent of the choice of x_0 provided $|x_0| > 1$. The levels of optimality $\{s_i\}_0^n$ satisfy $|T_n(s_i)| = 1$, $i = 0, 1, \cdots, n$, where the polynomial $T_n(x)$ is characterized by the property that (apart from a multiplicative constant) it minimizes the quantity $\sup_{-1 \leqq x \leqq 1} |x^n + \sum_{i=0}^{n-1} a_i x^i|$. Thus $T_n(x)$ in this special case is, of course, the well known Tchebycheff polynomial of the first kind.

In a recent paper, Kiefer and Wolfowitz (1965) extended this result by replacing the ordinary powers $1, x, \cdots, x^n$ by a T-system of functions $f_0, f_1, \cdots, f_n$ (see below). Subject to suitable restrictions on $f_0, f_1, \cdots, f_n$ the best approximation of f_n by polynomials $\sum_{i=0}^{n-1} a_i f_i$ attains its maximum absolute value at exactly $n + 1$ points $\{s_\nu^*\}_0^n$. The principal theorem of Kiefer and Wolfowitz (1965) provides a characterization of the vectors $\mathbf{c} = (c_0, \cdots, c_n)$ with the property that the optimum design for estimating the linear form $(\mathbf{c}, \boldsymbol{\theta})$ concentrates on the fixed set $\{s_i^*\}_0^n$.

In Theorem 3.1 below we state that part of Kiefer and Wolfowitz (1965) which is a direct extension of the Hoel and Levine result (see Corollary 3.1).

A system of continuous functions $\{f_i\}_0^n$ defined on an interval $[a, b]$ is called a *Tchebycheff system* or a T-system on $[a, b]$ provided every real linear combination $\sum_{i=0}^n a_i f_i(x)$ $(\sum_{i=0}^n a_i^2 > 0)$ has at most n distinct zeros on $[a, b]$. It is readily ascertained that this definition is equivalent to requiring that the determinants

$$U\begin{pmatrix} 0, & \cdots, & n \\ x_0, & \cdots, & x_n \end{pmatrix} = \begin{vmatrix} f_0(x_0) & f_0(x_1) & \cdots & f_0(x_n) \\ f_1(x_0) & f_1(x_1) & \cdots & f_1(x_n) \\ \vdots & \vdots & & \vdots \\ f_n(x_0) & f_n(x_1) & \cdots & f_n(x_n) \end{vmatrix} \tag{3.1}$$

are of one strict sign provided $a \leqq x_0 < x_1 < \cdots < x_n \leqq b$. For definiteness we shall suppose that the sign in (3.1) is positive.

The classical prototype of a T-system consists of the power functions $f_i(x) = x^i$, $i = 0, 1, \cdots, n$. In this case the determinant (3.1) reduces to the classical Vandermonde determinant. The "polynomials" $\sum_{i=0}^n a_i f_i(x)$ for a general T-system share many of the properties of ordinary polynomials $\sum_{i=0}^n a_i x^i$. Tchebycheff systems, in fact, play an important role in many domains of mathematics, notably the theory of approximations, methods of interpolation, generalized moment problems, numerical analysis, oscillation properties of eigenfunctions of Sturm-Liouville problems, generalized convexity, etc. A detailed discussion of many of these aspects of T-systems is presented in Karlin and Studden (1966).

The following important property of T-systems will be used in Theorem 3.1 below [see Karlin and Studden (1966), Theorem II.10.1]. If $\{f_i\}_0^n$ is a T-system on $[-1, 1]$ then there exists a unique polynomial $\underline{u}(x) = \sum_{i=0}^n a_i^* f_i(x)$ satisfying the properties:

(i) $|\underline{u}(x)| \leqq 1, x \varepsilon [-1, 1]$ and

(ii) there exist $n + 1$ points $-1 \leqq s_0 < s_1 < \cdots < s_n \leqq 1$ such that $\underline{u}(s_{n-i}) = (-1)^i, i = 0, 1, \cdots, n$.

Let $\mathcal{C}$ denote the set of vectors $\mathbf{c} = (c_0, c_1, \cdots, c_n)$ for which

$$\begin{vmatrix} f_0(x_1) & \cdots & f_0(x_n) & c_0 \\ f_1(x_1) & \cdots & f_1(x_n) & c_1 \\ \vdots & & \vdots & \vdots \\ f_n(x_1) & \cdots & f_n(x_n) & c_n \end{vmatrix} \neq 0 \tag{3.2}$$

for all $\{x_i\}_1^n$ satisfying $-1 \leqq x_1 < x_2 < \cdots < x_n \leqq 1$. Note that the determinants (3.2) necessarily maintain a constant sign. In the following it will be convenient to associate with each $\mathbf{c} \varepsilon \mathcal{C}$ an abstract point z and to set $f_i(z) = c_i$, $i = 0, 1, \cdots, n$. For any polynomial $u = \sum_{i=0}^n a_i f_i$ we may write $u(z) = \sum_{i=0}^n a_i f_i(z) = \sum_{i=0}^n a_i c_i$.

THEOREM 3.1. *Let $\{f_i\}_0^n$ be a T-system on $[-1, 1]$ and let u and $\mathcal{C}$ be defined as above.*

(a) *If $\mathbf{c} \varepsilon \mathcal{C}$ and ξ denotes a probability measure on $[-1, 1]$ then*

$$d(\mathbf{c}, \xi) \geqq \underline{u}^2(z), \quad (f_i(z) = c_i, i = 0, \cdots, n). \tag{3.3}$$

(b) *Let $s_0^* < s_1^* < \cdots < s_n^*$ be $n + 1$ points in the set $B = \{x \mid \underline{u}^2(x) = 1\}$ such that the values $\underline{u}(s_0^*), \cdots, \underline{u}(s_n^*)$ alternate in sign (the existence of at least one such set of $n + 1$ points is guaranteed by property* (ii) *of the polynomial $\underline{u}$). Let $L_\nu(x) = \sum_{i=0}^n a_i^{(\nu)} f_i(x)$ denote the Lagrange interpolation polynomials for the points $\{s_i^*\}_0^n$,* i.e., *$L_\nu(s_j^*) = \delta_{\nu j}, \nu, j = 0, 1, \cdots, n$. Then equality occurs in* (3.3) *when $\xi = \xi^*$ concentrates mass $p_i^* = |L_i(z)|/\sum_{j=0}^n |L_j(z)|$ at $s_i^*, i = 0, 1, \cdots, n$.*

(c) *If $\sum_{i=0}^n a_i f_i(x) \equiv 1, x \varepsilon [-1, 1]$ for some set of coefficients $\{a_i\}_0^n$ then B consists of exactly $n + 1$ points. Furthermore $s_0 = -1, s_n = +1$, and equality occurs in* (3.3) *if and only if ξ^* is defined as in part* (b).

PROOF. We first prove part (b) by appropriately applying the result of Theorem 2.2. Since the optimal designs for estimating the vectors $\mathbf{c}$ and $-\mathbf{c}$ are the same we may assume, without loss of generality, *that the determinants in* (3.2) *are positive.*

From the definition of a T-system it is readily seen that every polynomial is uniquely determined by its values at $n + 1$ distinct points. In this case the Lagrange polynomials $L_\nu(x)$ are well defined. Moreover for any polynomial $u(x) = \sum_{i=0}^n a_i f_i(x)$ we have $u(x) = \sum_{\nu=0}^n u(s_\nu^*) L_\nu(x)$ and hence

$$u(z) = \sum_{\nu=0}^n u(s_\nu^*) L_\nu(z).$$

In particular if we set $u = f_j$ then $f_j(z) = \sum_{\nu=0}^n f_j(s_\nu^*) L_\nu(z)$.

The polynomials $L_\nu(x)$ admit the explicit expression

$$L_\nu(x) = U\begin{pmatrix} 0, & \cdots, & \nu - 1, & \nu, & \nu + 1, & \cdots, & n \\ s_0^*, & \cdots, & s_{\nu-1}^*, & x, & s_{\nu+1}^*, & \cdots, & s_n^* \end{pmatrix} \Big/ U\begin{pmatrix} 0, & \cdots, & n \\ s_0^*, & \cdots, & s_n^* \end{pmatrix}$$

(see (3.1) regarding the notation) so that $L_\nu(z) = (-1)^{n-\nu}|L_\nu(z)|$, $\nu = 0, 1, \cdots, n$. Therefore, if we set $\epsilon_\nu = (-1)^{n-\nu}$ we have

$$\sum_{\nu=0}^n \epsilon_\nu |L_\nu(z)| f_j(s_\nu^*) = f_j(z), \qquad j = 0, 1, \cdots, n,$$

or equivalently

$$\sum_{\nu=0}^n \epsilon_\nu p_\nu^* \mathbf{f}(s_\nu^*) = \beta \mathbf{f}(z) = \beta \mathbf{c}$$

where $p_\nu^* = |L_\nu(z)| / \sum_{\nu=0}^n |L_\nu(z)|$, $\nu = 0, 1, \cdots, n$, and $\beta = [\sum_{\nu=0}^n |L_\nu(z)|]^{-1}$. Thus the first requirement of Theorem 2.2 holds.

We now show that $\beta\mathbf{c}$, for $\beta^{-1} = \sum_{k=0}^n |L_k(z)|$, lies on the boundary of $\mathcal{R}$. Let $\mathbf{a}^*$ denote the vector of coefficients of the polynomial $\underline{u}(z)$. Then $(\beta\mathbf{c}, \mathbf{a}^*) = \beta\underline{u}(z)$. Since the values s_ν^* were specified so that $\underline{u}(s_0^*)$, $\underline{u}(s_1^*)$, $\cdots$, $\underline{u}(s_n^*)$ have absolute value one and alternate in sign it follows that

$$[\underline{u}(z)]^2 = [\sum_\nu \underline{u}(s_\nu) L_\nu(z)]^2 = (\sum_{\nu=0}^n |L_\nu(z)|)^2.$$

Thus

$$\beta^2 = [\underline{u}(z)]^{-2}. \tag{3.4}$$

Consulting property (ii) of the polynomial $\underline{u}$ we find, in particular, that $\underline{u}(z) > 0$. It follows that

$$(\beta\mathbf{c}, \mathbf{a}^*) = \beta\underline{u}(z) = 1. \tag{3.5}$$

Moreover by property (i), we infer that $(\mathbf{f}(x), \mathbf{a}^*) = \underline{u}(x) \leqq 1$ and $(-\mathbf{f}(x), \mathbf{a}^*) = -\underline{u}(x) \leqq 1$ so that

$$(\mathbf{y}, \mathbf{a}^*) \leqq 1 \qquad \text{for all} \qquad \mathbf{y} \in \mathcal{R}. \tag{3.6}$$

The relations (3.5) and (3.6) imply that $\beta\mathbf{c}$ is a boundary point of $\mathcal{R}$. The demonstration of part (b) is hereby complete.

To prove part (a) we need simply observe that the minimum value of the left side of (3.3) is $v_0 = \beta^{-2} = \underline{u}^2(z)$ (see Theorem 2.2).

We now consider part (c). If $\sum_{i=0}^n a_i f_i(x) \equiv 1$, $x \in [-1, 1]$ then, according to properties (i) and (ii) on $\underline{u}$, we infer that the polynomials $1 - \underline{u}(x)$ and $1 + \underline{u}(x)$ are non-negative on $[-1, 1]$ and each exhibits n zeros with the convention that zeros in the open interval $(-1, 1)$ are counted twice. In this case the set B consists of exactly $n + 1$ points including $s_0^* = -1$ and $s_n^* = +1$ [see Theorem I.4.2 of Karlin and Studden (1966)].

Let ξ be an arbitrary probability measure on $[-1, 1]$ and assume that the spectrum of ξ includes points other than the $n + 1$ points of B. Since $\underline{u}^2(x) < 1$ at such points and $\underline{u}^2(x) \leqq 1$ for all $x \in [-1, 1]$ we find that

$$d(\mathbf{c}, \xi) = \sup_{\mathbf{d}} [(\mathbf{c}, \mathbf{d})^2/(\mathbf{d}, \mathbf{M}(\xi)\mathbf{d})] = \sup_{\mathbf{d}} [(\sum d_i f_i(z))^2/\int (\sum d_i f_i(x))^2 \xi(dx)]$$
$$\geqq \underline{u}^2(z)/\int \underline{u}(x)\xi(dx) > \underline{u}^2(z).$$

Therefore any ξ attaining equality in (3.3) necessarily concentrates on the $n + 1$ points $-1 = s_0^* < s_1^* < \cdots < s_n^* = 1$. Let ξ concentrate mass p_ν at s_ν^*, $\nu = 0, 1, \cdots, n$, respectively. If ξ is optimal then for some $\epsilon_\nu = \pm 1$,

$$\sum_{\nu=0}^{n} \epsilon_\nu p_\nu \mathbf{f}(s_\nu^*) = \beta \mathbf{f}(z), \qquad \beta = [\sum |L_\nu(z)|]^{-1}. \tag{3.7}$$

Regarding (3.7) as a system of equations in the unknown $\epsilon_\nu p_\nu$ and noticing that the corresponding determinant, namely det $\|f_i(s_\nu^*)\|$ is non-zero we infer that the solution in unique and hence $p_\nu = |L_\nu(z)|/\sum_{\nu=0}^{n} |L_\nu(z)|$ and $\epsilon_\nu = (-1)^{n-\nu}$, $\nu = 0, 1, \cdots, n$. Therefore, the optimal weights are as described in part (b).

COROLLARY 3.1. Hoel and Levine (1964). *Let $f_i(x) = x^i$, $i = 0, 1, \cdots, n$. If $|x_0| > 1$ and ξ denotes a probability measure on $[-1, 1]$ then*

$$d(x_0, \xi) \geqq T_n^2(x_0) \tag{3.8}$$

where $T_n(x)$ denotes the n*th Tchebycheff polynomial of the first kind. Moreover if*

$$s_\nu = -\cos(\nu\pi/n), \qquad \nu = 0, 1, \cdots, n, \tag{3.9}$$

and $L_\nu(x)$, $\nu = 0, 1, \cdots, n$, are the Lagrange interpolation polynomials of degree n for the points $\{s_\nu\}_0^n$ then equality occurs in (3.8) if and only if ξ concentrates mass $p_\nu^0 = |L_\nu(x_0)|/\sum_{j=0}^{n} |L_j(x_0)|$ at the points s_ν, $\nu = 0, 1, \cdots, n$.

4. Minimax designs. A classical criterion for selecting an optimal design is to choose ξ so as to maximize the determinant of the information matrix $\mathbf{M}(\xi)$. As noted previously this procedure is equivalent to minimizing the determinant of the covariance matrix of the best linear unbiased estimates of the parameters $\theta_0, \theta_1, \cdots, \theta_n$.

Another possible criterion for optimality may be to minimize the maximum over $x \varepsilon \mathfrak{X}$ of the variance $N^{-1} d(x, \xi)$ of the best linear unbiased estimate of the regression function $(\mathbf{f}(x), \boldsymbol{\theta}) = \sum_{i=0}^{n} \theta_i f_i(x)$.

It has been shown [see Kiefer and Wolfowitz (1960)] that these two criteria are equivalent.

The following proof of this theorem is new and utilizes game theoretic arguments which permit generalizations to be discussed in Section 6.

THEOREM 4.1. Kiefer-Wolfowitz (1960). (*Equivalence Theorem*). *The conditions*

(i) ξ^* *maximizes* $|\mathbf{M}(\xi)|$,
(ii) ξ^* *minimizes* $\sup_x d(x, \xi)$,

(iii) $\sup_x d(x, \xi^*) = n + 1$
are equivalent. The set Γ *of all* ξ^* *satisfying these conditions is convex and closed and* $\mathbf{M}(\xi^*)$ *is the same for all* $\xi^* \varepsilon \Gamma$.

PROOF. We first note that for any ξ for which $|\mathbf{M}(\xi)| = 0$ there exists some x_0 for which $\mathbf{f}(x_0)$ is not estimable, and hence $d(x_0, \xi) = \infty$ by definition. We may therefore restrict consideration to those ξ for which $|\mathbf{M}(\xi)| > 0$. In this case $d(x, \xi) = (\mathbf{f}(x), \mathbf{M}^{-1}(\xi)\mathbf{f}(x))$.

Consider the game with kernel $K_\epsilon(\xi, \eta) = \operatorname{tr} \mathbf{M}^{-1}(\eta)\mathbf{M}(\xi)$ where ξ and η range over the set

$$\Xi_\epsilon = \{\xi \mid \text{all eigenvalues of } \mathbf{M}(\xi) \text{ are } \geqq \epsilon\}.$$

The kernel $K_\epsilon(\xi, \eta)$ is linear and thus, *a fortiori*, concave in ξ. Using the fact that $\mathbf{M}^{-1}(\eta)$ as a matrix function is convex, i.e., the matrix $\alpha\mathbf{M}^{-1}(\eta_1) + (1 - \alpha)\mathbf{M}^{-1}(\eta_2) - [\alpha\mathbf{M}(\eta_1) + (1 - \alpha)\mathbf{M}(\eta_2)]^{-1}$ is positive definite for $0 < \alpha < 1$ and $\mathbf{M}(\eta_1) \neq \mathbf{M}(\eta_2)$, it is readily seen that $K_\epsilon(\xi, \eta)$ is convex in η. Since $K_\epsilon(\xi, \eta)$ is continuous in ξ and η and Ξ_ϵ is compact each of the games $(\Xi_\epsilon, \Xi_\epsilon, K_\epsilon)$ has a determined value v_ϵ and optimal strategies for both Players.

Clearly for each ξ $\inf_\eta \operatorname{tr} \mathbf{M}^{-1}(\eta)\mathbf{M}(\xi) \leqq n + 1$ since we can take $\eta = \xi$. Therefore

$$\text{(4.1)} \qquad \sup_\xi \inf_\eta \operatorname{tr} \mathbf{M}^{-1}(\eta)\mathbf{M}(\xi) \leqq n + 1$$

where the infinum and supremum are evaluated over Ξ_ϵ. Invoking the arithmetic-geometric mean inequality,[6] we obtain

$$\text{(4.2)} \qquad \operatorname{tr} \mathbf{M}^{-1}(\eta)\mathbf{M}(\xi) \geqq (n + 1)|\mathbf{M}(\xi)|^{(n+1)^{-1}}/|\mathbf{M}(\eta)|^{(n+1)^{-1}}$$

and equality occurs if and only if $\mathbf{M}(\xi)$ is proportional to $\mathbf{M}(\eta)$. Since $\prod_{i=0}^n \lambda_i(\xi) = |\mathbf{M}(\xi)|$ where $\lambda_i(\xi)$ are the eigenvalues of $\mathbf{M}(\xi)$ and

$$\max_i \lambda_i(\xi) = \sup_{\mathbf{v}\neq 0} [(\mathbf{v}, \mathbf{M}(\xi)\mathbf{v})/(\mathbf{v}, \mathbf{v})] \leqq \int (\textstyle\sum_{i=0}^n v_i f_i(x))^2 \xi(dx),$$

we see that the matrices $\mathbf{M}(\xi_0)$ attaining $\sup_\xi |\mathbf{M}(\xi)|$ (which is finite) necessarily have eigenvalues exceeding some ϵ_0.

For $\epsilon \leqq \epsilon_0$ it follows from (4.2) that

$$\text{(4.3)} \qquad \sup_{\xi\varepsilon\Xi_\epsilon} \inf_{\eta\varepsilon\Xi_\epsilon} \operatorname{tr} \mathbf{M}^{-1}(\eta)\mathbf{M}(\xi) \geqq n + 1.$$

Comparing (4.1) and (4.3) we see that the value of the game $(\Xi_\epsilon, \Xi_\epsilon, K_\epsilon)$, is $v_\epsilon = n + 1$ independent of $\epsilon \leqq \epsilon_0$.

Let $\Xi_0(\epsilon)$ and $\Theta_0(\epsilon)$ be the class of optimal strategies for Players I and II respectively. From (4.2) we infer that $\Theta_0(\epsilon) \subset \Gamma \subset \Xi_0(\epsilon)$ where $\Gamma = \{\xi \mid |\mathbf{M}(\xi)| = \sup_\eta |\mathbf{M}(\eta)|\}$

Now let $\xi_0 \varepsilon \Gamma \subset \Xi_0(\epsilon)$ and $\eta_0 \varepsilon \Theta_0(\epsilon)$. The optimality of ξ_0 and η_0 requires the relation $\operatorname{tr} \mathbf{M}^{-1}(\eta_0)\mathbf{M}(\xi_0) = n + 1$. Moreover, since ξ_0 and η_0 are in Γ, $|\mathbf{M}(\eta_0)| =$

[6] In the form $(\operatorname{tr} \mathbf{P})/n + 1 \geqq |\mathbf{P}|^{1/(n+1)}$ for $\mathbf{P}$ positive semi-definite with equality if and only if P is a multiple of the identity matrix.

$|\mathbf{M}(\xi_0)|$ so that from (4.2)

$$(4.4)\qquad \begin{aligned} n+1 &= \operatorname{tr} \mathbf{M}^{-1}(\eta_0)\mathbf{M}(\xi_0) \\ &\geqq (n+1)|\mathbf{M}(\xi_0)|^{(n+1)^{-1}}/|\mathbf{M}(\eta_0)|^{(n+1)^{-1}} = n+1. \end{aligned}$$

But equality occurs in (4.4) only when $\mathbf{M}(\xi_0)$ is proportional to $\mathbf{M}(\eta_0)$ and we may conclude that $\mathbf{M}(\xi_0) = \mathbf{M}(\eta_0)$. Keeping ξ_0 fixed, we deduce that the matrices $\mathbf{M}(\eta_0)$, $\eta_0 \,\varepsilon\, \Theta_0(\epsilon)$, all coincide. Furthermore, ξ_0 was specified arbitrarily in Γ so that $\Theta_0(\epsilon) = \Gamma$. The set $\Gamma = \Theta_0(\epsilon)$ is therefore convex and closed.

We can now prove that conditions (i), (ii) and (iii) define the same set. Since $\sup_x d(x, \xi^*) = \sup_\xi \operatorname{tr} \mathbf{M}^{-1}(\xi^*)\mathbf{M}(\xi) \geqq n+1$ it follows that (ii) and (iii) determine the set $\Theta_0(\epsilon)$. Furthermore (i) is precisely the set Γ so that the desired result is established.

5. Maximization of certain determinants. In this section our objective is to determine, for special choices of the functions $f_0, f_1, \cdots, f_n$, those designs ξ_0 which maximize $|\mathbf{M}(\xi)|$ or equivalently minimize $\sup_x d(x, \xi)$. In carrying this out we find that the maximization of $|\mathbf{M}(\xi)|$ is more amenable to direct analysis. We cannot always display the solution in closed form but some characterizations are available.

Stating the problem more explicitly we wish to determine the measures ξ_0 which maximize the determinant of the matrix $\mathbf{M}(\xi) = \|m_{ij}(\xi)\|$ where $m_{ij}(\xi) = \int_{\mathfrak{X}} f_i(x)f_j(x)\xi(dx)$, $i, j = 0, 1, \cdots, n$, for special choices of $f_0, f_1, \cdots, f_n$.

Suppose the measure ξ concentrates mass $p_0, p_1, \cdots, p_n (\sum_{j=0}^n p_j = 1)$ at $n+1$ distinct points $x_0, \cdots, x_n$, respectively. For this choice of ξ the elements of $\mathbf{M}(\xi)$ become

$$(5.1)\qquad m_{ij} = \textstyle\sum_{l=0}^n p_l f_i(x_l) f_j(x_l).$$

Writing $a_{il} = p_l f_i(x_l)$ and $b_{lj} = f_j(x_l)$ it follows that $|\mathbf{M}(\xi)| = |\mathbf{A}||\mathbf{B}|$ so that the determinant of $\mathbf{M}(\xi)$ takes the special form

$$(5.2)\qquad \{\textstyle\prod_{l=0}^n p_l\}\{\det \|f_i(x_j)\|_{i,j=0}^n\}^2.$$

We observe from (5.2) that the two sets of variables $x_0, x_1, \cdots, x_n$ and $p_0, p_1, \cdots, p_n$ separate. The maximization of (5.2) can therefore be performed in two stages. An elementary calculation reveals that the maximum of $\prod_{l=0}^n p_l$ under the conditions $p_l \geqq 0$, $\sum_{l=0}^n p_l = 1$ is achieved for $p_l = (n+1)^{-1}$, $l = 0, 1, \cdots, n$. In general the maximization of det $\|f_i(x_j)\|$ or even the simpler task of characterizing the solution appears very formidable.

If the functions $f_0, \cdots, f_n$ are specified to be

$$(5.3)\qquad f_i(x) = x^i w^{\frac{1}{2}}(x), \qquad i = 0, 1, \cdots, n, \qquad x \,\varepsilon\, [a, b],$$

then (5.2) becomes

$$(5.4)\qquad (\textstyle\prod_{l=0}^n p_l)\prod_{l=0}^n w(x_l)\prod_{0\leqq i<j\leqq n}(x_i - x_j)^2.$$

Remark 5.1. The above considerations suggest a number of pertinent inquiries. For example, it would be of interest to determine conditions on the functions $f_0, f_1, \cdots, f_n$ that guarantee that the maximum of $|\mathbf{M}(\xi)|$ will be attained for a measure concentrating at $n + 1$ points. In this case $|\mathbf{M}(\xi)|$ simplifies to (5.2).

A second important area of inquiry concerns the problem of characterizing those weight functions $w(x)$ for the functions $f_i(x) = x^i w^{\frac{1}{2}}(x)$, $i = 0, 1, \cdots, n$, whose associated maximizing measure ξ contains $n + 1$ jumps. A partial solution is indicated in Theorem 5.2 below. It would also be of interest to characterize the points $x_0, x_1, \cdots, x_n$ which produce the maximum in either (5.2) or (5.4).

We now turn to the maximization of the determinant $|\mathbf{M}(\xi)|$ when the functions f_i, $i = 0, 1, \cdots, n$, are of the form (5.3) and $w(x)$ is a classical weight function.

THEOREM 5.1. *Let* $f_i(x) = x^i w^{\frac{1}{2}}(x)$, $i = 0, \cdots, n$, *and let* $w(x)$ *be one of the following weight functions*;

(i) $w(x) \equiv 1$, $[a, b] = [-1, 1]$;
(ii) $w(x) = (1 - x)^{\alpha+1}(1 + x)^{\beta+1}$, $[a, b] = [-1, 1]$, $\alpha > -1$, $\beta > -1$;
(iii) $w(x) = e^{-x}$, $[a, b] = [0, \infty]$;
(iv) $w(x) = x^{\alpha+1}e^{-x}$, $[a, b] = [0, \infty]$, $\alpha > -1$;
(v) $w(x) = e^{-x^2}$, $[a, b] = [-\infty, \infty]$.

Then the determinant $|\mathbf{M}(\xi)|$ *is uniquely maximized by the measure* ξ_0 *concentrating equal mass* $(n + 1)^{-1}$ *at the zeros of the polynomials*

(i) $(1 - x^2)P_n'(x)$ *where* P_n *is the* nth *Legendre polynomial*;
(ii) $P_{n+1}^{(\alpha,\beta)}(x)$, *the* $n + 1$th *Jacobi polynomial*;
(iii) $xL_n^{(1)}(x)$ *where* $L_n^{(\alpha)}(x)$ *is the* nth *Laguerre polynomial of parameter* α;
(iv) $L_{n+1}^{(\alpha)}(x)$;
(v) *the Hermite polynomial* $H_{n+1}(x)$.

PROOF. Since the analysis in each case is similar we shall consider only part (iv). We first set $D(\xi) = |\mathbf{M}(\xi)|$ and note that $0 < \sup_\xi D(\xi) = D_0 < \infty$. Invoking the classical Helly selection theorem we obtain a sequence $\{\xi_n\}$ satisfying $\lim_{n\to\infty} D(\xi_n) = D_0$ and ξ_n converges weakly to ξ_0 (which, *a priori*, may not be a bona fide probability measure, i.e., ξ_0 may have total measure less than one). Since each of the functions $x^{\nu+\alpha+1}e^{-x}$, $\nu = 0, 1, \cdots, n$, vanishes at infinity, it follows that $D(\xi_0) = D_0$ so that the maximum is attained for ξ_0. We claim that ξ_0 cannot exhibit mass at the origin nor can the total measure $\xi_0(\infty)$ be less than 1. If we assume to the contrary that $\xi_0(\infty) < 1$ then the measure $\xi_0(dx)/\xi_0(\infty)$ produces a determinant of value exceeding D_0 which is manifestly absurd. The possibility that ξ_0 concentrates mass at zero is precluded by analogous reasoning.

Now let $L_\nu(\xi) = \int_0^\infty x^{\nu+\alpha+1}e^{-x}\, d\xi$, $\nu = 0, 1, \cdots, 2n$. Because $\max_\xi D(\xi)$ is attained for $\xi = \xi_0$ it follows that the maximum of $L_{2n}(\xi)$, extended over those probability measures whose first $2n$ moments are $L_\nu(\xi_0)$, $\nu = 0, 1, \cdots, 2n - 1$, is achieved for $\xi = \xi_0$. In this case the point $(L_0(\xi_0), L_1(\xi_0), \cdots, L_{2n}(\xi_0))$ is a boundary point of the convex set $M = \{(L_0(\xi), \cdots, L_{2n}(\xi))\}$ where ξ varies over the set of all probability measures on $[0, \infty)$. Consequently there exists a supporting hyperplane to M at $(L_0(\xi_0), \cdots, L_{2n}(\xi_0))$ i.e., there exist real constants $a_0, \cdots, a_{2n}$ and d with $\sum_0^{2n} a_\nu^2 > 0$ such that $\sum_0^{2n} a_i L_i(\xi) \leqq d$, for all ξ and $\sum_0^{2n} a_i L_i(\xi_0) = d$. Equivalently, $\int_0^\infty (\sum_{i=0}^{2n} a_i x^{i+\alpha+1} e^{-x} - d)\xi(dx) \leqq 0$ for all ξ, with equality for ξ_0. This implies that $\sum_{i=0}^{2n} a_i x^{i+\alpha+1} e^{-x} \leqq d$ for all $x \varepsilon [0, \infty)$ and equality occurs at all points of increase of ξ_0. We now prove that equality holds for at most $n + 1$ points exceeding zero. If equality occurs at $n + 2$ points then the function $g(x) = e^{-x} x^{\alpha+1} P_{2n}(x) - d$, where $P_{2n}(x) = \sum_0^{2n} a_i x^i$, possesses at least $2n + 4$ zeros counting multiplicities. On the basis of Rolle's theorem we may conclude that $g'(x) = e^{-x} x^\alpha[-xP_{2n} + xP'_{2n} + (\alpha + 1)P_{2n}]$ has at least $2n + 3$ zeros on $(0, \infty)$, counting multiplicities.

This requires $-xP_{2n}(x) + xP'_{2n}(x) + (\alpha + 1)P_{2n}(x) \equiv 0$ and hence $P_{2n}(x) \equiv 0$ which contradicts the fact that $\sum_{i=0}^{2n} a_i^2 > 0$. Since $D(\xi_0) > 0$ it is clear that ξ_0 necessarily admits at least $n + 1$ points of increase. The above analysis proves that the measure ξ_0 consists of exactly $n + 1$ mass points $x_0, \cdots, x_n$, all greater than zero, with corresponding weights $p_0, p_1, \cdots, p_n$, $\sum_{i=0}^n p_i = 1$.

We may therefore write

$$L_\nu(\xi_0) = \sum_{i=0}^n x_i^{\nu+\alpha+1} e^{-x_i} p_i . \tag{5.5}$$

Referring to (5.4) it follows that $D(\xi_0)$ reduces to

$$D(\xi_0) = \left(\prod_0^n p_i\right) \exp\left[-\sum x_i\right]\left(\prod_0^n x_i^{\alpha+1}\right)\prod_{0\leqq i<j\leqq n} (x_i - x_j)^2. \tag{5.6}$$

We may now proceed along the lines of the proof of Theorems 6.7.1–6.7.3 in Szegö [1959]. In this case we may differentiate $\log D(\xi_0)(\prod_0^n p_i)^{-1}$ and verify that $f(x) = \prod_0^n (x - x_i)$ satisfies the differential equation which determines the Laguerre polynomial $L_{n+1}^{(\alpha)}(x)$ up to a constant factor.

As remarked earlier the product $\prod_0^n p_i$ under the constraints $p_i \geqq 0$, $\sum_{i=0}^n p_i = 1$ is maximized when $p_i = (n + 1)^{-1}$. This completes the proof of the theorem.

Remark 5.2. Theorem 5.1 was discovered by Schoenberg (1959). Schoenberg used rather involved variational arguments while the above method of proof is geometric and leads to a more general theorem (Theorem 5.2) given below. The explicit values of $|\mathbf{M}(\xi_0)|$ in each case can be determined by referring to the discriminant functions of the classical polynomials, see Szegö (1959), p. 141.

Some of the preceding arguments pertaining to the number of points in the spectrum of a measure ξ_0 which maximizes $|\mathbf{M}(\xi)|$ are directly applicable for the case $f_i(x) = [w(x)]^{\frac{1}{2}} x^i$, $(i = 0, 1, \cdots, n, a \leqq x \leqq b)$ where $w(x)$ is a general weight function.

We shall assume that $w(x)$ is continuous on the interval (a, b), which may be finite or infinite, and that if $a = -\infty$ $(b = +\infty)$ then the limit of $w(x)x^{2n}$ as $x \to a$ $(x \to b)$ is finite. We let $\mu_\nu = \int_a^b w(x)x^\nu \xi(dx)$, $\nu = 0, 1, \cdots, 2n$, and denote, as previously, the determinant of $\|\mu_{i+j}\|_{i,j=0}^n$ by $D(\xi)$. We shall assume that $w(x) > 0$ for at least $n + 1$ points; otherwise $D(\xi) \equiv 0$.

If $D(\xi_0) = \max_\eta D(\eta)$ we may deduce using the supporting hyperplane argument, as in the proof of Theorem 5.1, that an upper bound on the number of points in the spectrum of ξ_0 is the maximum number of points for which equality holds in the relation $w(x)\sum_{i=0}^{2n} a_i x^i \leqq d$, $x \varepsilon [a, b]$, $\sum_{i=0}^{2n} a_i^2 > 0$. Using this fact we prove the following theorem.

THEOREM 5.2. *The maximum of $D(\eta)$ with respect to the set of all probability measures is attained by a measure ξ_0 which concentrates at exactly $n + 1$ points if any of the following conditions hold*:

(i) *the system $\{1, w(x), xw(x), \cdots, x^{2n}w(x)\}$ is a T-system on $[a, b]$, (see Section 3).*

(ii) *$w(x) = [P(x)]^{-1}$, where $P(x)$ is a polynomial, positive on $[a, b]$ and $P^{(2n+1)}(x)$ has no zeros on the open interval (a, b), (e.g. if $P(x)$ has only real zeros either all less than a or all greater than b).*

(iii) *$w(x)$ can be approximated uniformly by weight functions of the type considered in (ii).*

(iv) *$w(x) = [P(x)]^{-1}$, where $P(x)$ is a polynomial, positive on $[a, b]$, of degree at most $2n$.*

REMARK 5.3. Case (iii) applies if $a = 0, b = \infty$ and $w(x)$ is the Laplace transform of a Pólya frequency function on $[0, \infty)$, see Schoenberg (1951).

PROOF. Suppose that (i) holds. From the remarks preceding the statement of the theorem there exist real a_i, $i = 0, 1, \cdots, 2n$, $(\sum a_i^2 > 0)$ such that $w(x)\sum_0^{2n} a_i x^i \leqq d$ for some real d and equality occurs on the spectrum of ξ_0. If there exist at least $n + 2$ points $x_1 < x_2 < \cdots < x_{n+2}$ in the spectrum of ξ_0, then three alternatives can arise: $x_i \varepsilon (a, b)$ for all i; $n + 1$ of the x_i's lie in (a, b) and one is located at an endpoint a or b; $x_2, \cdots, x_{n+1}$ are interior to (a, b), $x_1 = a$ and $x_{n+2} = b$. For each of the three possibilities the "polynomial" $w(x)\sum_0^{2n} a_i x^i - d$ has at least $2n + 2$ zeros where zeros in the open interval (a, b) are counted twice. This is impossible by Theorem I.4.2 of Karlin and Studden (1966).

If condition (ii) is postulated, then as above, the function $g(x) = \sum_0^{2n} a_i x^i - d \cdot P(x)$ admits at least $2n + 2$ zeros in the interval $[a, b]$ where zeros at the endpoints a and b are counted once. If $d = 0$ this is impossible while if $d \neq 0$ then the function $g^{(2n+1)}(x) = -d \cdot P^{(2n+1)}(x)$ possesses a zero in (a, b) contradicting the hypothesis.

The third part follows by a straightforward limiting argument.

Part (iv) is handled in a slightly different manner. Let

$$M = \{(\mu_0(\xi), \cdots, \mu_{2n}(\xi))\} \qquad (\mu_\nu = \int_a^b w(x)x^\nu \xi(dx))$$

denote the moment space obtained by varying ξ over all probability measures on $[a, b]$ and let $\mathfrak{M}$ denote the closed convex cone generated by M. Since $P(x)$ is a polynomial of degree at most $2n$ we may write $P(x) = \sum_{i=0}^{2n} b_i x^i$. It follows that M is the section of the cone $\mathfrak{M}$ consisting of the points $(c_0, \cdots, c_{2n}) \varepsilon \mathfrak{M}$ satisfying the normalization condition $\sum_{i=0}^{2n} b_i c_i = 1$. Now if ξ_0 maximizes the determinant $D(\xi)$ then $(\mu_0(\xi_0), \cdots, \mu_{2n}(\xi_0))$ is necessarily a boundary point of M. Since M is a section of $\mathfrak{M}$ the point $(\mu_0(\xi_0), \cdots, \mu_{2n}(\xi_0))$ must also be a boundary point of the cone $\mathfrak{M}$. In this case there exist real constants $a_0, a_1, \cdots, a_{2n}$ ($\sum_{i=0}^{2n} a_i^2 > 0$) such that

$$\sum_{i=0}^{2n} a_i x^i / w(x) \leqq 0, \qquad a \leqq x \leqq b, \tag{5.7}$$

and equality holds for x in the spectrum of ξ_0. But $|\mathbf{M}(\xi_0)| > 0$ implies that the spectrum of ξ_0 contains at least $n + 1$ points. However $w(x) > 0$ for $x \varepsilon [a, b]$ so that equality can hold in (5.7) for at most $n + 1$ points which includes the two endpoints a and b. Therefore the spectrum of ξ_0 contains exactly $n + 1$ points.

6. Extensions of Theorem 4.1. In Theorem 4.1 it was shown that any measure ξ maximizing $|\mathbf{M}(\xi)|$ also minimizes $\max_x d(x, \xi)$ and conversely. The significance of this theorem was the identification of the optimal designs associated with seemingly distinct statistical criteria. The matrix $\mathbf{M}(\xi)$ is the information matrix associated with an experiment ξ while $N^{-1} d(x, \xi)$ corresponds, for fixed ξ, to the variance of the best linear unbiased estimate of the regression function $\sum_{i=0}^{n} \theta_i f_i(x)$.

The functions $\mathbf{M}(\xi)$ and $d(x, \xi)$ are relevant for the problem of estimating the full set of $n + 1$ parameters $\theta_0, \cdots, \theta_n$. In considering the estimation of only the first $s + 1$ of the $n + 1$ parameters we introduce two corresponding functions. When $\mathbf{M}(\xi)$ is non-singular, the analog of $|\mathbf{M}(\xi)|$ will be

$$|\mathbf{M}_s^*(\xi)| = |\mathbf{M}_1(\xi) - \mathbf{M}_2'(\xi)\mathbf{M}_3^{-1}(\xi)\mathbf{M}_2(\xi)| \tag{6.1}$$

where

$$\mathbf{M}(\xi) = \begin{Vmatrix} \mathbf{M}_1(\xi) & \mathbf{M}_2'(\xi) \\ \mathbf{M}_2(\xi) & \mathbf{M}_3(\xi) \end{Vmatrix} \qquad (\mathbf{M}_1(\xi) \text{ is } s + 1 \times s + 1). \tag{6.2}$$

The analog of $d(x, \xi)$ is

$$\begin{aligned} d_s(x, \xi) &= \operatorname{tr} \mathbf{P}(\xi)(\mathbf{f}^{(1)}(x) - \mathbf{D}'(\xi)\mathbf{f}^{(2)}(x))(\mathbf{f}^{(1)}(x) - \mathbf{D}'(\xi)\mathbf{f}^{(2)}(x))' \\ &= (\mathbf{f}^{(1)}(x) - \mathbf{D}'(\xi)\mathbf{f}^{(2)}(x), \mathbf{P}(\xi)(\mathbf{f}^{(1)}(x) - \mathbf{D}'(\xi)\mathbf{f}^{(2)}(x))) \end{aligned} \tag{6.3}$$

where

$$\mathbf{f}^{(1)}(x) = (f_0(x), \cdots, f_s(x)), \qquad \mathbf{f}^{(2)}(x) = (f_{s+1}(x), \cdots, f_n(x)),$$
$$\mathbf{P}(\xi) = [\mathbf{M}_s^*(\xi)]^{-1} \quad \text{and} \quad \mathbf{D}(\xi) = \mathbf{M}_3^{-1}(\xi)\mathbf{M}_2(\xi).$$

Our objective in this section is to extend the equivalence theorem (Theorem 4.1) to show that a measure ξ maximizes $|\mathbf{M}_s^*(\xi)|$ if and only if it minimizes $\max_x d_s(x, \xi)$.

We first wish to extend the definition of $\mathbf{M}_s^*(\xi)$ to the case where $\mathbf{M}_3(\xi)$ is singular. For this purpose we need the following elementary facts that are consequences of the hypothesis that $\mathbf{M}(\xi)$ in (6.2) is positive semi-definite.

LEMMA 6.1. *Range* $\mathbf{M}_2(\xi) \subset$ *Range* $\mathbf{M}_3(\xi)$.

PROOF. The statement of the lemma is equivalent to the relation

$$[\text{Range } \mathbf{M}_2(\xi)]^\perp \supset [\text{Range } \mathbf{M}_3(\xi)]^\perp$$

($\perp$ denotes orthogonal complement). But for any matrix $\mathbf{C}$ the null space $\mathfrak{N}(\mathbf{C}')$ of $\mathbf{C}'$ ($\mathbf{C}'$ = the transpose of $\mathbf{C}$) coincides with [Range $(\mathbf{C})]^{\perp}$. Therefore, we need to show that $\mathfrak{N}(\mathbf{M}_2'(\xi)) \supset \mathfrak{N}(\mathbf{M}_3'(\xi))$.

Since $\mathbf{M}(\xi)$ is non-negative definite we know that if $\mathbf{x} = \{\epsilon\mathbf{x}_1, \mathbf{x}_2\}$ ($\mathbf{x}_1$ is a vector of $s + 1$ coordinates and $\mathbf{x}_2$ is a vector of $n - s$ coordinates), then for all real ϵ,

$$(\mathbf{x}, \mathbf{M}(\xi)\mathbf{x}) = \epsilon^2(\mathbf{x}_1, \mathbf{M}_1(\xi)\mathbf{x}_1) + 2\epsilon(\mathbf{x}_1, \mathbf{M}_2'(\xi)\mathbf{x}_2) + (\mathbf{x}_2, \mathbf{M}_3(\xi)\mathbf{x}_2) \geqq 0.$$

Let $\mathbf{x}_2$ be contained in $\mathfrak{N}(\mathbf{M}_3(\xi)) = \mathfrak{N}(\mathbf{M}_3'(\xi))$. Then $(\mathbf{x}_2, \mathbf{M}_3(\xi)\mathbf{x}_2) = 0$ and it follows from the preceding inequality that $(\mathbf{x}_1, \mathbf{M}_2'(\xi)\mathbf{x}_2) = 0$ for all $\mathbf{x}_1$. Therefore $\mathbf{x}_2 \in \mathfrak{N}(\mathbf{M}_2'(\xi))$ which implies $\mathfrak{N}(\mathbf{M}_2'(\xi)) \supset \mathfrak{N}(\mathbf{M}_3'(\xi))$.

COROLLARY 6.1. *A solution* $\mathbf{X}$ *of order* $(n - s) \times (s + 1)$ *of the matrix equation* $\mathbf{M}_3(\xi)\mathbf{X} = \mathbf{M}_2(\xi)$ *always exists.*

LEMMA 6.2. *The matrix* $\mathbf{X}'\mathbf{M}_3(\xi)\mathbf{X}$ *is independent of* $\mathbf{X}$ *provided* $\mathbf{X}$ *is a solution of* $\mathbf{M}_3(\xi)\mathbf{X} = \mathbf{M}_2(\xi)$.

PROOF. If $\mathbf{M}_3(\xi)\mathbf{D} = \mathbf{0}$ then trivially $(\mathbf{X}' + \mathbf{D}')\mathbf{M}_3(\mathbf{X} + \mathbf{D}) = \mathbf{X}'\mathbf{M}\mathbf{X}$.

We now set $\mathbf{M}_s^*(\xi) = \mathbf{M}_1(\xi) - \mathbf{X}(\xi)'\mathbf{M}_3(\xi)\mathbf{X}(\xi)$ where $\mathbf{X}(\xi)$ is a solution of $\mathbf{M}_3(\xi)\mathbf{X}(\xi) = \mathbf{M}_2(\xi)$. By virtue of Lemma 6.2 $\mathbf{M}_s^*(\xi)$ is well defined independent of the particular solution used. We also define $d_s(x, \xi)$ as in (6.3) with $\mathbf{D}(\xi)$ replaced by $\mathbf{X}(\xi)$ provided $\mathbf{M}_s^*(\xi)$ is nonsingular.

The quantity $\mathbf{M}_s^*(\xi)$ enjoys the following properties:

LEMMA 6.3. (i) $\mathbf{M}_s^* = \mathbf{M}_s^*(\xi)$ *is non-negative definite.*

(ii) $\mathbf{M}_s^*(\xi) = \mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi) \leqq \mathbf{M}_1(\xi)$.

PROOF. (i) If $\mathbf{D}$ is an arbitrary $(n - s) \times (s + 1)$ matrix and $\mathbf{I}$ represents the $(s + 1) \times (s + 1)$ identity matrix then

$$\|\mathbf{I}, -\mathbf{D}'\| \begin{Vmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{Vmatrix} \begin{Vmatrix} \mathbf{I} \\ -\mathbf{D} \end{Vmatrix} = \mathbf{M}_1 - \mathbf{M}_2'\mathbf{D} - \mathbf{D}'\mathbf{M}_2 + \mathbf{D}'\mathbf{M}_3\mathbf{D}$$

is manifestly non-negative definite. In particular let $\mathbf{D}$ satisfy $\mathbf{M}_3\mathbf{D} = \mathbf{M}_2$. Then the above matrix becomes

$$\mathbf{M}_1 - \mathbf{M}_2'\mathbf{D} - \mathbf{D}'\mathbf{M}_2 + \mathbf{D}'\mathbf{M}_3\mathbf{D} = \mathbf{M}_1 - \mathbf{M}_2'\mathbf{D} = \mathbf{M}_1 - \mathbf{D}'\mathbf{M}_3\mathbf{D} = \mathbf{M}_s^*.$$

Thus $\mathbf{M}_s^* = \mathbf{M}_s^*(\xi)$ is always non-negative definite.

(ii) This result follows immediately since $\mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)$ is clearly non-negative definite.

The following two elementary lemmas will also be needed.

LEMMA 6.4. *If* $\mathbf{X}(\xi)$ *is a solution of* $\mathbf{M}_3(\xi)\mathbf{X}(\xi) = \mathbf{M}_2(\xi)$ *then the matrix*

$$\begin{aligned}\mathbf{X}'\mathbf{M}_3\mathbf{X} - \mathbf{X}'\mathbf{M}_3\mathbf{D} - \mathbf{D}'\mathbf{M}_3\mathbf{X} + \mathbf{D}'\mathbf{M}_3\mathbf{D} \\ = \mathbf{X}'\mathbf{M}_3\mathbf{X} - \mathbf{M}_2'\mathbf{D} - \mathbf{D}'\mathbf{M}_2 + \mathbf{D}'\mathbf{M}_3\mathbf{D}\end{aligned}$$

is non-negative definite. Moreover this matrix is identically zero if and only if $\mathbf{D}$ *also satisfies* $\mathbf{M}_3\mathbf{D} = \mathbf{M}_2$.

PROOF. The above matrix coincides with

$$(\mathbf{X}' - \mathbf{D}')\mathbf{M}_3(\mathbf{X} - \mathbf{D}) \tag{6.4}$$

which is clearly non-negative definite and certainly vanishes when $\mathbf{D}$ satisfies $\mathbf{M}_3\mathbf{D} = \mathbf{M}_2$. Moreover if $\mathbf{M}_3(\mathbf{X} - \mathbf{D}) \neq \mathbf{0}$ then $\mathbf{A} = \mathbf{M}_3^{\frac{1}{2}}(\mathbf{X} - \mathbf{D}) \neq \mathbf{0}$ which brings (6.4) into the form $\mathbf{A}\mathbf{A}'$ with $\mathbf{A} \neq \mathbf{0}$. This obviously vanishes if and only if $\mathbf{A} = \mathbf{0}$ or what is the same if and only if $\mathbf{D}$ satisfies $\mathbf{M}_3\mathbf{D} = \mathbf{M}_2$.

LEMMA 6.5. *If* $\mathbf{C}$ *is an arbitrary non-negative definite* $(s + 1) \times (s + 1)$ *matrix then*

$$\inf_{|\mathbf{P}|=1} \operatorname{tr} \mathbf{P}\mathbf{C} = (s + 1)|\mathbf{C}|^{(s+1)^{-1}} \tag{6.5}$$

where the infimum is extended over the set of positive definite matrices $\mathbf{P}$ (*order* $(s + 1) \times (s + 1)$) *of determinant one. In the case that* $|\mathbf{C}| > 0$, equality occurs in (6.5) *if and only if* $\mathbf{P}$ *is proportional to* $\mathbf{C}^{-1}$.

This is another form of the arithmetic mean geometric mean inequality (see footnote, 6, Section 4).

We are now prepared to analyze the functions $\mathbf{M}_s^*(\xi)$ and $d_s(x, \xi)$. Since the functions f_i, $i = 0, \cdots, n$, are continuous and $\mathfrak{X}$ is compact it follows from Lemma 6.3(ii) that

$$0 < \sup_\xi |\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)| = a < \infty. \tag{6.6}$$

Let $\{\mathbf{M}(\xi_k)\}$ be a sequence of information matrices for which

$$|\mathbf{M}_1(\xi_k) - \mathbf{X}'(\xi_k)\mathbf{M}_3(\xi_k)\mathbf{X}(\xi_k)| \to a$$

and without loss of generality we take ξ_k converging weak* to some ξ_0. Then

$$\lim_{k\to\infty} \mathbf{M}(\xi_k) = \mathbf{M}(\xi_0). \tag{6.7}$$

It is not clear at present that the matrices $\mathbf{X}(\xi_k)$ converge; however, it will be shown below that $a = |\mathbf{M}_s^*(\xi_0)| = |\mathbf{M}_1(\xi_0) - \mathbf{X}'(\xi_0)\mathbf{M}_3(\xi_0)\mathbf{X}(\xi_0)|$.

We introduce the function

$$\begin{aligned} \phi(\mathbf{P}, \mathbf{D}; \xi) &= \operatorname{tr} \mathbf{P}(\mathbf{M}_1 - \mathbf{D}'\mathbf{M}_2 - \mathbf{M}_2'\mathbf{D} + \mathbf{D}'\mathbf{M}_3\mathbf{D}) \\ &= \int \operatorname{tr} \mathbf{P}(\mathbf{f}^{(1)} - \mathbf{D}'\mathbf{f}^{(2)})(\mathbf{f}^{(1)} - \mathbf{D}'\mathbf{f}^{(2)})'\, \xi(dx) \end{aligned} \tag{6.8}$$

where $\mathbf{D}$ is an arbitrary $r \times s + 1$ $(r = n - s)$ matrix and $\mathbf{P}$ is positive definite of determinant value a^{-1}. Obviously

$$\begin{aligned} \phi(\mathbf{P}, \mathbf{D}; \xi) &= \operatorname{tr} \mathbf{P}(\mathbf{M}_1(\xi) - \mathbf{D}'\mathbf{M}_2(\xi) - \mathbf{M}_2'(\xi)\mathbf{D} + \mathbf{D}'\mathbf{M}_3(\xi)\mathbf{D}) \\ &= \operatorname{tr} \mathbf{P}(\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi) + \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi) \\ &\quad - \mathbf{D}'\mathbf{M}_2(\xi) - \mathbf{M}_2'(\xi)\mathbf{D} + \mathbf{D}'\mathbf{M}_3(\xi)\mathbf{D}) \end{aligned}$$

so that from Lemma 6.4

$$\phi(\mathbf{P}, \mathbf{D}; \xi) \geqq \operatorname{tr} \mathbf{P}(\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)) \tag{6.9}$$

and equality holds in (6.9) if and only if $\mathbf{D}$ satisfies $\mathbf{M}_3(\xi)\mathbf{D} = \mathbf{M}_2(\xi)$. Therefore

$$\inf_{\mathbf{P},\mathbf{D},|\mathbf{P}|=a^{-1}} \phi(\mathbf{P}, \mathbf{D}; \xi) = \inf_{\mathbf{P},|\mathbf{P}|=a^{-1}} \operatorname{tr} \mathbf{P}(\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)). \tag{6.10}$$

Now appealing to Lemma 6.5 we deduce that

$$(6.11)\quad \begin{aligned}&\inf_{\mathbf{P},|\mathbf{P}|=a^{-1}} \operatorname{tr} \mathbf{P}(\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi))\\ &\qquad = (s+1)|\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)|^{(s+1)^{-1}}/a^{(s+1)^{-1}}\end{aligned}$$

where a is defined in (6.6). Combining (6.10) and (6.11) we obtain for all probability measure ξ that

$$(6.12)\quad \begin{aligned}&\inf_{\mathbf{P},\mathbf{D},|\mathbf{P}|=a^{-1}} \phi(\mathbf{P}, \mathbf{D}; \xi)\\ &\qquad = (s+1)|\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)|^{(s+1)^{-1}}/a^{(s+1)^{-1}}\end{aligned}$$

and

$$(6.13)\quad \sup_\xi \inf_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi) = s + 1.$$

If ξ is such that $|\mathbf{M}_s^*(\xi)| = |\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)| > 0$ then Lemma 6.5 informs us that equality occurs in (6.12) if and only if $\mathbf{D}$ satisfies $\mathbf{M}_3(\xi)\mathbf{D} = \mathbf{M}_2(\xi)$ and $\mathbf{P}$ is proportional to $(\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi))^{-1}$; the proportionality constant is determined by the condition $|\mathbf{P}| = a^{-1}$.

Finally, since

$$\phi(\mathbf{P}, \mathbf{D}; \xi_k) \geqq (s+1)|\mathbf{M}_1(\xi_k) - \mathbf{X}'(\xi_k)\mathbf{M}_3(\xi_k)\mathbf{X}(\xi_k)|^{(s+1)^{-1}}/a^{(s+1)^{-1}}$$

the limiting properties of the sequence $\{\xi_k\}$ imply that

$$(6.14)\quad \phi(\mathbf{P}, \mathbf{D}; \xi_0) \geqq s + 1.$$

Therefore

$$(s+1)|\mathbf{M}_1(\xi_0) - \mathbf{X}'(\xi_0)\mathbf{M}_3(\xi_0)\mathbf{X}(\xi_0)|^{(s+1)^{-1}}/a^{(s+1)^{-1}} = \inf_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi_0) \geqq s + 1$$

and hence

$$(6.15)\quad |\mathbf{M}_1(\xi_0) - \mathbf{X}'(\xi_0)\mathbf{M}_3(\xi_0)\mathbf{X}(\xi_0)| = a.$$

The above analysis will help in proving the following theorem.

THEOREM 6.1. *Let $\mathcal{P}$ denote the set of all positive definite matrices (of order $(s+1) \times (s+1)$) normalized so that $|\mathbf{P}| = a^{-1}$ and let $\mathfrak{D}$ comprise the set of all real matrices of order $(n-s) \times (s+1)$. The kernel defined in* (6.8) *satisfies*

$$\max_\xi \min_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi) = s + 1 = \min_{\mathbf{P},\mathbf{D}} \max_\xi \phi(\mathbf{P}, \mathbf{D}; \xi).$$

Furthermore, ξ_0 determined in (6.7) *satisfies*

$$\phi(\mathbf{P}, \mathbf{D}; \xi_0) \geqq s + 1 \quad \textit{for all} \quad (\mathbf{P}, \mathbf{D}) \in \mathcal{P} \times \mathfrak{D}.$$

If $(\mathbf{P}_0, \mathbf{D}_0) \in \mathcal{P} \times \mathfrak{D}$ fulfills

$$\phi(\mathbf{P}_0, \mathbf{D}_0; \xi) \leqq s + 1 \quad \textit{for all probability measures } \xi \textit{ on } \mathfrak{X}$$

then $\mathbf{D}_0$ satisfies $\mathbf{M}_3(\xi_0)\mathbf{D}_0 = \mathbf{M}_2(\xi_0)$ and $\mathbf{P}_0$ is a constant multiple of $[\mathbf{M}_1(\xi_0) - \mathbf{D}_0'\mathbf{M}_3(\xi_0)\mathbf{D}_0]^{-1}$.

PROOF. We consider the sequence of games

$$\Psi_N(\mathbf{Q}, \mathbf{E}; \xi) = \operatorname{tr}(\mathbf{Q}'\mathbf{M}_1(\xi)\mathbf{Q} + \mathbf{Q}'\mathbf{M}_2'(\xi)\mathbf{E} + \mathbf{E}'\mathbf{M}_2(\xi)\mathbf{Q} + \mathbf{E}'\mathbf{M}_3(\xi)\mathbf{E})$$

$$= \operatorname{tr}\left(\|\mathbf{Q}', \mathbf{E}'\| \begin{Vmatrix} \mathbf{M}_1(\xi) & \mathbf{M}_2'(\xi) \\ \mathbf{M}_2(\xi) & \mathbf{M}_3(\xi) \end{Vmatrix} \begin{Vmatrix} \mathbf{Q} \\ \mathbf{E} \end{Vmatrix}\right)$$

where $\mathbf{E}$ denotes an arbitrary $(n - s) \times (s + 1)$ real matrix whose elements are uniformly bounded by N and $\mathbf{Q}$ stands for a general positive definite matrix $((s + 1) \times (s + 1))$ normalized so that $|\mathbf{Q}| \geqq a^{-\frac{1}{2}}$ with smallest eigenvalue $\geqq N^{-1}$. The ξ variable, as previously, traverses the set Ξ of all probability measures on $\mathfrak{X}$.

We denote the strategy space of the minimizing player by $\mathcal{Q}_N \times \mathcal{E}_N$. In view of the classical inequality

$$|\mathbf{Q}_1 + \mathbf{Q}_2|^{(s+1)^{-1}} \geqq |\mathbf{Q}_1|^{(s+1)^{-1}} + |\mathbf{Q}_2|^{(s+1)^{-1}}$$

when $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are positive semi-definite it is readily established that $\mathcal{Q}_N$ is convex and compact and obviously the same holds for $\mathcal{E}_N$. Clear Ξ is convex and compact since $\mathfrak{X}$ is compact. Moreover, we note the fact that the kernel $\Psi_N(\mathbf{Q}, \mathbf{E}; \xi)$ is convex with respect to $(\mathbf{Q}, \mathbf{E})$ and linear in ξ. The fundamental theorem of the theory of games affirms the existence of $(\mathbf{Q}_N, \mathbf{E}_N) \varepsilon\, \mathcal{Q}_N \times \mathcal{E}_N$ and $\xi_N \varepsilon\, \Xi$ fulfilling the relations

$$(6.16) \qquad \Psi_N(\mathbf{Q}_N, \mathbf{E}_N; \xi) \leqq v_N \qquad \text{for all} \quad \xi \,\varepsilon\, \Xi$$

and

$$(6.17) \qquad \Psi_N(\mathbf{Q}, \mathbf{E}; \xi_N) \geqq v_N \qquad \text{for all} \quad (\mathbf{Q}, \mathbf{E}) \,\varepsilon\, \mathcal{Q}_N \times \mathcal{E}_N$$

for some constant v_N. Clearly v_N decreases as N increases since only $\mathcal{Q}_N \times \mathcal{E}_N$ enlarges. This, shows in particular, that v_N is uniformly bounded.

Now choosing ξ so that $\mathbf{M}_3(\xi)$ is non-singular we infer from (6.16) that the elements of $\mathbf{Q}_N$ and $\mathbf{E}_N$ are uniformly bounded. Selecting a pair of limit matrices $\mathbf{Q}_0$ and $\mathbf{E}_0$ from $\{\mathbf{Q}_N\}$ and $\{\mathbf{E}_N\}$ respectively and a weak* limit measure $\bar{\xi}$ from ξ_N and executing the obvious limit procedure in (6.16) and (6.17) leads to the inequalities

$$(6.18) \qquad \Psi(\mathbf{Q}_0, \mathbf{E}_0; \xi) \leqq v, \qquad \text{for all} \quad \xi \,\varepsilon\, \Xi$$

and

$$(6.19) \qquad \Psi(\mathbf{Q}, \mathbf{E}; \bar{\xi}) \geqq v \qquad \text{for all} \quad (\mathbf{Q}, \mathbf{E}) \,\varepsilon\, \mathcal{Q} \times \mathcal{E}$$

where $v = \lim_{N\to\infty} v_N$. Obviously $\mathbf{Q}_0$ is a positive definite matrix of order $(s + 1) \times (s + 1)$ fulfilling the condition $|\mathbf{Q}_0| \geqq a^{-\frac{1}{2}}$. Since $|\mathbf{Q}_0| \geqq a^{-\frac{1}{2}}$ we can introduce the matrix $\mathbf{D}_0 = -\mathbf{E}_0 \mathbf{Q}_0^{-1}$ and $\mathbf{D} = -\mathbf{E}\mathbf{Q}^{-1}$ generally for $(\mathbf{Q}, \mathbf{E})$ in $\mathcal{Q} \times \mathcal{E}$. Then (6.18) and (6.19) become

$$(6.20) \qquad \phi(\mathbf{P}_0, \mathbf{D}_0; \xi) \leqq v \qquad \text{for all} \quad \xi \,\varepsilon\, \Xi$$

and

$$(6.21) \qquad \phi(\mathbf{P}, \mathbf{D}; \tilde{\xi}) \geqq v \quad \text{for all} \quad (\mathbf{P}, \mathbf{D}) \in \mathcal{P} \times \mathfrak{D}.$$

We see directly from the definition that $\phi(\lambda\mathbf{P}, \mathbf{D}; \xi) = \lambda\phi(\mathbf{P}, \mathbf{D}; \xi)$ for $\lambda > 0$. Therefore, we may suppose that $\mathbf{P}_0$ satisfies the normalization condition $|\mathbf{P}_0| = a^{-1}$; otherwise replace $\mathbf{P}_0$ by $\lambda\mathbf{P}_0$ for appropriate λ $(0 < \lambda < 1)$ and relations (6.20) are manifestly preserved.

The inequalities (6.20) and (6.21) imply the identity

$$(6.22) \qquad \max_\xi \min_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi) = v = \min_{\mathbf{P},\mathbf{D}} \max_\xi \phi(\mathbf{P}, \mathbf{D}; \xi).$$

Comparison of (6.22), (6.13) and (6.14) shows that $v = s + 1$ and that we can replace $\tilde{\xi}$ in (6.21) by ξ_0. Since $\mathbf{P}_0$, $\mathbf{D}_0$ minimizes $\phi(\mathbf{P}, \mathbf{D}; \xi_0)$ the characterization of $\mathbf{P}_0$ and $\mathbf{D}_0$ as stated in the theorem follows as indicated in the discussion of (6.14). The proof of the theorem is complete.

Using the above theorem we may deduce the following theorem.

THEOREM 6.2. Kiefer (1962). *Let $f_0, \cdots, f_n$ be linearly independent continuous functions on a compact space $\mathfrak{X}$. Then for $0 \leqq s \leqq n$ there exists real numbers a_{ij}, $0 \leqq i \leqq s$, $0 \leqq j \leqq n$, with the matrix $\|a_{ij}\|$, $(0 \leqq i, j \leqq s)$ non-singular and a probability measure ξ_0 on $\mathfrak{X}$ with finite spectrum such that*

(a) *the functions $g_i = \sum_{j=0}^{n} a_{ij} f_j$, $0 \leqq i \leqq s$, are orthonormal with respect to ξ_0 and are orthogonal to f_j for $s < j \leqq n$.*

(b) $\max_x \sum_{i=0}^{s} g_i^2(x) = \int \sum_{i=0}^{s} g_i^2(x)\xi_0(dx) = s + 1.$

PROOF. By Lemma 2.1 we can choose ξ_0 satisfying (6.14) with finite spectrum. Let $\mathbf{D}_0$ and $\mathbf{P}_0$ be determined as in Theorem 6.1. We define

$$(g_0, \cdots, g_s) = \mathbf{g} = \mathbf{P}_0^{\frac{1}{2}}\{\mathbf{f}^{(1)} - \mathbf{D}_0'\mathbf{f}^{(2)}\}.$$

Since $\mathbf{P}_0[(\mathbf{M}_1(\xi_0) - \mathbf{D}_0'\mathbf{M}_2(\xi_0) - \mathbf{M}_2'(\xi_0)\mathbf{D}_0 + \mathbf{D}_0'\mathbf{M}_3(\xi_0)\mathbf{D}_0] = \mathbf{I}$ (the identity matrix) we conclude that the functions $g_0, \cdots, g_s$ are orthonormal with respect to ξ_0. Moreover, since $\phi(\mathbf{P}_0, \mathbf{D}_0 + \epsilon\mathbf{D}, \xi_0) \geqq \phi(\mathbf{P}_0, \mathbf{D}_0; \xi_0)$ for all $\mathbf{D} \in \mathfrak{D}$ and arbitrary real ϵ it follows that

$$-2\epsilon \int \operatorname{tr} \mathbf{g}(x)\mathbf{f}^{(2)}(x)'\mathbf{D}\mathbf{P}_0^{\frac{1}{2}}\xi_0(dx) + \epsilon^2 \operatorname{tr} \mathbf{P}_0\mathbf{D}'\mathbf{M}_3(\xi_0)\mathbf{D} \geqq 0.$$

However this is possible only if the coefficient of ϵ is zero. Now choosing $\mathbf{D}$ so that $\mathbf{D}\mathbf{P}_0^{\frac{1}{2}}$ is the $(n - s) \times (s + 1)$ matrix with zeros everywhere except in the jth row and ith column, we have $\int g_i f_{s+j}\, d\xi_0 = 0, 0 \leqq i \leqq s, 0 < j \leqq n - s$. This completes the proof of (a). To prove (b) note that

$$\begin{aligned} s + 1 = \phi(\mathbf{P}_0, \mathbf{D}_0; \xi_0) &= \sup_\xi \phi(\mathbf{P}_0, \mathbf{D}_0; \xi) \\ &= \sup_\xi \int (\operatorname{tr} \mathbf{g}(x)\mathbf{g}'(x))\xi(dx) = \sup_x \textstyle\sum_{i=0}^{s} g_i^2(x) \end{aligned}$$

so that

$$\max_x \sum_{i=0}^{s} g_i^2(x) = s + 1 \qquad \text{and} \qquad \int \sum_{i=0}^{s} g_i^2(x)\xi_0(dx) = s + 1.$$

The proof of Theorem 6.2 is complete.

We now turn to the analog of Theorem 4.1. For any measure such that $|\mathbf{M}_s^*(\xi)| > 0$ we define (see (6.3))

(6.23) $$d_s(x, \xi) = \operatorname{tr} \mathbf{P}(\xi)(\mathbf{f}^{(1)}(x) - \mathbf{D}'(\xi)\mathbf{f}^{(2)}(x))(\mathbf{f}^{(1)}(x) - \mathbf{D}'(\xi)\mathbf{f}^{(2)}(x))'$$

where $\mathbf{P}(\xi) = [\mathbf{M}_s^*(\xi)]^{-1} = [\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)]^{-1}$ and $\mathbf{M}_3(\xi)\mathbf{X}(\xi) = \mathbf{M}_2(\xi)$.

In the case that $\mathbf{M}(\xi)$ is non-singular it can be verified by juggling the obvious relations between $\mathbf{M}_i$ and $\mathbf{M}^{(i)}$

$$\left(\mathbf{M}(\xi)^{-1} = \begin{Vmatrix}\mathbf{M}^{(1)}(\xi) & \mathbf{M}^{(2)}(\xi)' \\ \mathbf{M}^{(2)}(\xi) & \mathbf{M}^{(3)}(\xi)\end{Vmatrix}\right)$$

that $d_s(x, \xi)$ reduces to

$$d_s(x, \xi) = (\mathbf{f}(x), \mathbf{M}^{-1}(\xi)\mathbf{f}(x)) - (\mathbf{f}^{(2)}(x), \mathbf{M}_3^{-1}(\xi)\mathbf{f}^{(2)}(x)).$$

The three conditions corresponding to those given in Theorem 4.1 are

(6.24)
(i) ξ^* maximizes $|\mathbf{M}_s^*(\xi)| = |\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)|$;
(ii) ξ^* minimizes $\max_x d_s(x, \xi)$;
(iii) $\max_x d_s(x, \xi^*) = s + 1$.

The analog of Theorem 4.1 is the following:

THEOREM 6.2. *The set of measures* ξ^* *fulfilling* (i), (ii) *or* (iii) *of* (6.24) *coincide.*

PROOF. If ξ^* satisfies (ii) then

$$\begin{aligned}\max_x d_s(x, \xi^*) &= \min_\xi \max_x d_s(x, \xi) = \min_\xi \max_\eta \int d_s(x, \xi)\, d\eta(x) \\ &= \min_\xi \max_\eta \phi(\mathbf{P}(\xi), \mathbf{D}(\xi), \eta) \leqq \max_\eta \phi(\mathbf{P}_0, \mathbf{D}_0; \eta) \leqq s + 1.\end{aligned}$$

But since $\phi(\mathbf{P}(\xi), \mathbf{D}(\xi); \xi) = s + 1$ we see that ξ^* satisfies (iii). Condition (iii) clearly implies (ii) since always $\max_x d_s(x, \xi) \geqq s + 1$.

If $\max_x d_s(x, \xi^*) = s + 1$ then taking account of (6.12) we deduce that

$$s + 1 = \max_\xi \phi(\mathbf{P}(\xi^*), \mathbf{D}(\xi^*), \xi) = \max_\xi \lambda^{-1}\phi(\lambda\mathbf{P}(\xi^*), \mathbf{D}(\xi^*), \xi) \geqq (s + 1)/\lambda$$

where λ is such that $|\mathbf{P}(\xi^*)\lambda| = a^{-1}$ ($a = \sup_\xi |\mathbf{M}_s^*(\xi)|$). Therefore $|a\mathbf{P}(\xi^*)| \leqq 1$ or $a \leqq |\mathbf{M}_s^*(\xi^*)|$ and hence condition (iii) implies condition (i).

From Equation (6.11) any ξ^* which maximizes $|\mathbf{M}_s^*(\xi)|$ fulfills $\phi(\mathbf{P}, \mathbf{D}; \xi^*) \geqq s + 1$. However, in this case, the optimal minimizing pair $\mathbf{P}_0$ and $\mathbf{D}_0$ are characterized by the properties that $\mathbf{P}_0 = \mathbf{P}(\xi^*)$ and $\mathbf{D}_0$ satisfies $\mathbf{M}_3(\xi^*)\mathbf{D}_0 = \mathbf{M}_2(\xi^*)$ so that $\phi(\mathbf{P}(\xi^*), \mathbf{D}(\xi^*); \xi) \leqq s + 1$ for all ξ and hence $\max_x d_s(x, \xi^*) = s + 1$. This completes the proof.

REMARK 6.1. The expression $\phi(\mathbf{P}, \mathbf{D}; \xi)$ can be written as

$$\phi(\mathbf{P}, \mathbf{D}; \xi) = \operatorname{tr} \bar{\mathbf{P}}\mathbf{M}(\xi)\bar{\mathbf{P}}'$$

where $\bar{\mathbf{P}}$ is the $s + 1 \times n + 1$ matrix defined by $\bar{\mathbf{P}} = \|\mathbf{Q}, -\mathbf{Q}\mathbf{D}'\|$, $\mathbf{Q}$ being the positive square root of $\mathbf{P}$. When $\mathbf{M}(\xi)$ is non-singular we can express $\phi(\mathbf{P}, \mathbf{D}; \xi)$ in the form

(6.25) $$\phi(\mathbf{P}, \mathbf{D}; \xi) = \sup_{\mathbf{E}\neq 0} [\operatorname{tr}^2 \bar{\mathbf{P}}\mathbf{E}/\operatorname{tr} \mathbf{E}'\mathbf{M}(\xi)^{-1}\mathbf{E}]$$

where $\mathbf{E}$ is an arbitrary $n + 1 \times s + 1$ matrix. Equality is attained for $\mathbf{E} = \mathbf{M}(\xi)\bar{\mathbf{P}}'$. The proof of (6.25) can be based on the inequality $\text{tr}^2 \bar{\mathbf{P}}\mathbf{E} \leqq (\text{tr } \mathbf{E}'\mathbf{M}(\xi)^{-1}\mathbf{E})(\text{tr } \bar{\mathbf{P}}\mathbf{M}(\xi)\bar{\mathbf{P}}')$ which is easily derived by employing an appropriate form of Schwartz's inequality. In the present case we have the formula

$$\inf_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi) = \inf_{\mathbf{P},\mathbf{D}} \sup_{\mathbf{E}\neq 0} [\text{tr}^2 (\bar{\mathbf{P}}\mathbf{E})/\text{tr } \mathbf{E}'\mathbf{M}(\xi)^{-1}\mathbf{E}].$$

REMARK 6.2. Whenever $|\mathbf{M}(\xi)| > 0$, it is easy to show that

$$|\mathbf{M}_1(\xi) - \mathbf{X}'(\xi)\mathbf{M}_3(\xi)\mathbf{X}(\xi)| = |\mathbf{M}_1(\xi) - \mathbf{M}_2'(\xi)\mathbf{M}_3^{-1}(\xi)\mathbf{M}_2(\xi)|$$

reduces to $|\mathbf{M}(\xi)|\, |\mathbf{M}_3(\xi)|^{-1}$ so that (cf. (6.11))

$$\inf_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi) = [(s + 1)/a^{(s+1)^{-1}}](|\mathbf{M}(\xi)|/|\mathbf{M}_3(\xi)|)^{(s+1)^{-1}}.$$

REMARK 6.3. In the case $s = 0$, $\phi(\mathbf{P}, \mathbf{D}; \xi)$ becomes (apart from a fixed multiplication factor)

$$\int_{\mathfrak{X}} (f_0(x) - \textstyle\sum_{j=1}^{n} d_j f_j(x))^2 \xi(dx).$$

Therefore the task of evaluating $\inf_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi)$ and determining the minimizing $\mathbf{D}$ is equivalent to finding that linear combination f^* of $f_1, \cdots, f_n$ for which $f_0 - f^*$ has minimum $L_2(\xi)$ norm.

If we view the functions f_i as random variables we wish to find f^* which minimizes what is essentially the variance of $f_0 - f^*$. This linear combination is the same as the one which gives maximum correlation between f_0 and f^* where the square of the correlation is $(\int f_0 f^* \, d\xi)^2/\int f_0^2 \, d\xi \int (f^*)^2 \, d\xi$.

In the case $(f_0(x), \cdots, f_n(x)) = (x^n, \cdots, x, 1)$ for $x \in [0, 1]$, the result of Theorem 6.1 specializes to

$$\sup_\xi \inf_{\mathbf{d}} \int_0^1 (x^n - \textstyle\sum_{i=0}^{n-1} d_i x^i)^2 \xi(dx) = \inf_{\mathbf{d}} \sup_x |x^n - \textstyle\sum_{i=0}^{n-1} d_i x^i|^2$$

and the right side is attained when the polynomial is proportional to the nth Tchebycheff polynomial of the first kind.

In the general case we have the following: For a fixed ξ we define an inner product by $\langle \mathbf{h}, \mathbf{g} \rangle = \int (\mathbf{h}, \mathbf{P}(\xi)\mathbf{g}) \, d\xi$ where $\mathbf{h} = (h_0, \cdots, h_s)$ and $\mathbf{g} = (g_0, \cdots, g_s)$. Then

$$\begin{aligned} \phi(\mathbf{P}(\xi), \mathbf{D}; \xi) &= \int_{\mathfrak{X}} ((\mathbf{f}^{(1)} - \mathbf{D}'\mathbf{f}^{(2)}), \mathbf{P}(\xi)(\mathbf{f}^{(1)} - \mathbf{D}'\mathbf{f}^{(2)})) \, d\xi \\ &= \langle \mathbf{f}^{(1)} - \mathbf{D}'\mathbf{f}^{(2)}, \mathbf{f}^{(1)} - \mathbf{D}'\mathbf{f}^{(2)} \rangle. \end{aligned}$$

We easily deduce that

$$\begin{aligned} &\inf_\lambda \phi(\mathbf{P}(\xi), \lambda\mathbf{D}, \xi) \\ &\quad = \langle \mathbf{f}^{(1)}, \mathbf{f}^{(1)} \rangle [1 - \langle \mathbf{f}^{(1)}, \mathbf{D}'\mathbf{f}^{(2)} \rangle^2 / \langle \mathbf{f}^{(1)}, \mathbf{f}^{(1)} \rangle \langle \mathbf{D}'\mathbf{f}^{(2)}, \mathbf{D}'\mathbf{f}^{(2)} \rangle] \quad (\lambda \text{ real}) \end{aligned}$$

so that

$$\begin{aligned} \inf_{\mathbf{P},\mathbf{D}} \phi(\mathbf{P}, \mathbf{D}; \xi) &= \inf_{\mathbf{D}} \phi(\mathbf{P}(\xi), \mathbf{D}; \xi) = \inf_{\mathbf{D}} \langle \mathbf{f}^{(1)} - \mathbf{D}'\mathbf{f}^{(2)}, \mathbf{f}^{(1)} - \mathbf{D}'\mathbf{f}^{(2)} \rangle \\ &= \langle \mathbf{f}^{(1)}, \mathbf{f}^{(1)} \rangle \{1 - \sup_{\mathbf{D}} [\langle \mathbf{f}^{(1)}, \mathbf{D}'\mathbf{f}^{(2)} \rangle^2 / \langle \mathbf{f}^{(1)}, \mathbf{f}^{(1)} \rangle \langle \mathbf{D}'\mathbf{f}^{(2)}, \mathbf{D}'\mathbf{f}^{(2)} \rangle]\}. \end{aligned}$$

7. Admissible designs. A measure or design ξ is called *admissible* if there does not exist a design ξ^* such that $\mathbf{M}(\xi^*) \geqq \mathbf{M}(\xi)$ (the inequality signifies that $\mathbf{M}(\xi^*) - \mathbf{M}(\xi)$ is positive semi-definite and $\mathbf{M}(\xi^*) \neq \mathbf{M}(\xi)$).

The next theorem is stated without proof in Elfving (1959) in a less precise form. Two examples will be given at the culmination of the proof to show that the converse statements of both parts of the theorem are not valid.

THEOREM 7.1. *Let $S(\xi_0)$ denote the spectrum of ξ_0.*

(i) *If there exists a positive definite matrix* $\mathbf{T}$ *such that* $(\mathbf{f}(x), \mathbf{Tf}(x)) \leqq 1$ *for all $x \varepsilon \mathfrak{X}$ and equality holds for $x \varepsilon S(\xi_0)$ then ξ_0 is admissible.*

(ii) *If ξ_0 is admissible then there exists a non-negative matrix* $\mathbf{T}$ (*not necessarily positive definite*) *such that* $(\mathbf{f}(x), \mathbf{Tf}(x)) \leqq 1$ *for all $x \varepsilon \mathfrak{X}$ with equality occurring for $x \varepsilon S(\xi_0)$.*

PROOF. (i) Suppose that there exists a design ξ^* such that $(\mathbf{d}, \mathbf{M}(\xi^*)\mathbf{d}) \geqq (\mathbf{d}, \mathbf{M}(\xi_0)\mathbf{d})$ for all $\mathbf{d}$ or equivalently

$$\int (\mathbf{f}(x), \mathbf{d})^2 \xi^*(dx) \geqq \int (\mathbf{f}(x), \mathbf{d})^2 \xi_0(dx) \qquad \text{for all} \quad \mathbf{d}. \tag{7.1}$$

Since the matrix $\mathbf{T}$ is positive definite it may be represented in the form $\mathbf{T} = \sum_0^n \mathbf{t}_j\mathbf{t}_j'$ where $\mathbf{t}_0, \cdots, \mathbf{t}_n$ are $n + 1$ linearly independent vectors. Substituting $\mathbf{t}_j$ for $\mathbf{d}$ in (7.1) and summing over j yields

$$\int (\mathbf{f}(x), \mathbf{Tf}(x))\xi^*(dx) \geqq \int (\mathbf{f}(x), \mathbf{Tf}(x))\xi_0(dx). \tag{7.2}$$

Since $(\mathbf{f}(x), \mathbf{Tf}(x)) \leqq 1$ for all x and equality occurs for $x \varepsilon S(\xi_0)$, the relation (7.2) is an equality. In this event, equality necessarily holds in (7.1) for the choices $\mathbf{d} = \mathbf{t}_j$; $j = 0, 1, \cdots, n$. Therefore $(\mathbf{t}_j, \mathbf{M}(\xi^*)\mathbf{t}_j) = (\mathbf{t}_j, \mathbf{M}(\xi_0)\mathbf{t}_j)$, $j = 0, 1, \cdots, n$, and since $\mathbf{M}(\xi^*) - \mathbf{M}(\xi_0)$ is semi-definite it follows that $\mathbf{M}(\xi^*) = \mathbf{M}(\xi_0)$. Therefore ξ_0 is admissible.

(ii) Let ξ_0 be admissible. Consider the game with kernel $\phi(\xi, \mu) = \int (\mathbf{g}, (\mathbf{M}(\xi) - \mathbf{M}(\xi_0))\mathbf{g})\mu(d\mathbf{g})$ where Player I maximizes over the set of probability measures ξ on $\mathfrak{X}$ and Player II minimizes over the set of probability measures μ defined on the unit sphere $\{\mathbf{g} \mid \mathbf{g} \varepsilon E^{n+1}, \|\mathbf{g}\| = 1\}$. The kernel $\phi(\xi, \mu)$ is linear in both variables and the strategy spaces are compact so that both players have optimal strategies ξ^* and μ^*. Moreover, since ξ_0 is admissible the value of the game is manifestly equal to zero. Therefore

$$\int (\mathbf{g}, \mathbf{M}(\xi)\mathbf{g})\mu^*(d\mathbf{g}) \leqq \int (\mathbf{g}, \mathbf{M}(\xi_0)\mathbf{g})\mu^*(d\mathbf{g}) \qquad \text{for all} \quad \xi. \tag{7.3}$$

The right side of (7.3) is positive due to the fact that $\mathbf{M}(\xi)$ is positive definite for some ξ. Letting $m_0 = \int (\mathbf{g}, \mathbf{M}(\xi_0)\mathbf{g})\mu^*(d\mathbf{g})$ and $\mathbf{T} = m_0^{-1} \int \mathbf{g}\mathbf{g}'\mu^*(d\mathbf{g})$ we may rewrite (7.3) in the form

$$\int (\mathbf{f}(x), \mathbf{Tf}(x))\xi(dx) \leqq 1 \qquad \text{for all} \quad \xi \tag{7.4}$$

and equality occurs for $\xi = \xi_0$. In this case $(\mathbf{f}(x), \mathbf{Tf}(x)) \leqq 1$ for all $x \varepsilon \mathfrak{X}$ with equality occurring when $x \varepsilon S(\xi_0)$. This completes the proof of the theorem.

We now present two examples which show that the converse statements of parts (i) and (ii) in Theorem 7.1 are false.

Consider part (ii). We wish to demonstrate that the existence of a non-negative matrix T satisfying $(\mathbf{f}(x), \mathbf{Tf}(x)) \leqq 1$ with equality on the spectrum of ξ does not imply ipso facto that ξ is admissible. To this end, consider $f_0(x) = 1, f_1(x) = x$ and $[a, b] = [0, 1]$. If

$$\mathbf{T} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

then $(\mathbf{f}(x), \mathbf{Tf}(x)) \equiv 1$.

If the converse of part (ii) were true then every ξ would be admissible. However, this is manifestly not the case. In fact, let ξ_{t_0} denote the measure concentrating all of its mass at t_0 $(0 < t_0 < 1)$ and let ξ_0 concentrate mass t_0 and $1 - t_0$ at 1 and 0 respectively. Then

$$\mathbf{M}(\xi_{t_0}) = \begin{pmatrix} 1 & t_0 \\ t_0 & t_0^2 \end{pmatrix}, \quad \text{and} \quad \mathbf{M}(\xi_0) = \begin{pmatrix} 1 & t_0 \\ t_0 & t_0 \end{pmatrix}$$

so that $\mathbf{M}(\xi_0) \geqq \mathbf{M}(\xi_{t_0})$.

The following example shows that the converse assertion of part (i) is false. Let $\mathfrak{X} = [0, \pi/2] \cup \{\pi\}$ and define $\mathbf{f} = (f_0, f_1)$ by $f(\pi) = (1, 1)$ and $\mathbf{f}(\theta) = (f_0(\theta), f_1(\theta)) = (\cos\theta, -\sin\theta)$. It can easily be checked that the design concentrating exclusively at $\theta = 0$ is admissible. However, there exists no positive definite matrix $\mathbf{T}$ satisfying $(\mathbf{f}(\theta), \mathbf{Tf}(\theta)) \leqq 1$ for all $\theta \in \mathfrak{X}$ and $(\mathbf{f}(0), \mathbf{Tf}(0)) = 1$.

We will next establish that the concept of admissibility of a design ξ_0 is essentially a property of the spectrum of ξ_0 (see Corollary 7.2 below).

THEOREM 7.2. *Let ξ_0 be an admissible design. Then every design ξ of the form $\xi(dx) = \varphi(x)\xi_0(dx)$ where $0 \leqq \varphi(x) \leqq K$ (K is a constant) is also admissible.*

PROOF. Suppose to the contrary that ξ is inadmissible. Then there exists a design ξ^* such that

$$\int (\mathbf{f}(x), \mathbf{d})^2 \xi^*(dx) \geqq \int (\mathbf{f}(x), \mathbf{d})^2 \xi(dx) \quad \text{for all } \mathbf{d} \tag{7.5}$$

and strict inequality holds for some $\mathbf{d}$.

Let $\xi_\eta = (1 - \eta)\xi + \eta\xi^*$ $(0 \leqq \eta)$. Then

$$\int (\mathbf{f}(x), \mathbf{d})^2 (\xi_\eta - \xi)(dx) = \eta \int (\mathbf{f}(x), \mathbf{d})^2 (\xi^* - \xi)(dx) \geqq 0$$

with strict inequality for some $\mathbf{d}$. It follows that

$$\int (\mathbf{f}(x), \mathbf{d})^2 (\xi_\eta - \xi + \xi_0)(dx) \geqq \int (\mathbf{f}(x), \mathbf{d})^2 \xi_0(dx). \tag{7.6}$$

Now since $\varphi(x) \leqq K$ it follows that for $\eta < K^{-1}$ the measure $(\xi_\eta - \xi + \xi_0)(dx) = \eta\xi^*(dx) + \xi_0(dx) - \eta\varphi(x)\xi_0(dx)$ is non-negative. Moreover

$$\int (\xi_\eta - \xi + \xi_0)(dx) = 1.$$

Equation (7.6) shows that the design ξ_0 is also inadmissible. This contradiction implies that the design ξ is admissible.

COROLLARY 7.2. *If a design $\xi_0 = \{x_i\,; p_i\}_1^r$ $(p_i > 0, \sum_{i=1}^r p_i = 1)$ is admissible then the design $\xi = \{x_i\,; q_i\}_1^r (q_i \geqq 0, \sum_{i=1}^r q_i = 1)$ is also admissible.*

For the case of ordinary polynomial regression on a finite interval $[a, b]$ the class of admissible designs have been completely characterized by Kiefer (1959).

The following theorem, which makes use of Theorem 7.1, part (i), incorporates an extension of his result.

THEOREM 7.3. *Let* $\mathfrak{X} = [a, b]$ *and* $f_i(x) = [w(x)]^{\frac{1}{2}}x^i$, $i = 0, 1, \cdots, n$, *where* $w^{-1}(x) = P_{2n-2}(x) = \sum_{i=0}^{2n-2} b_i x^i$ *is a polynomial, positive on* $(-\infty, \infty)$, *of degree at most* $2n - 2$. *Then a design* ξ_0 *is admissible if and only if the spectrum of* ξ_0 *contains at most* $n - 1$ *points in the open interval* (a, b).

PROOF. Suppose the spectrum of ξ_0 includes at most $n - 1$ points in (a, b). Then there exists a polynomial $Q_{n-1}(x)$ of exact degree $n - 1$ such that $(a - x)(b - x)Q_{n-1}^2(x)$ vanishes for x in the spectrum of ξ_0. Since $P_{2n-2}(x) > 0$ on $x \varepsilon [a, b]$ we specify ϵ sufficiently small and positive so that

$$P_{2n-2}(x) + \epsilon(a - x)(b - x)Q_{n-1}^2(x) > 0, \qquad x \varepsilon [a, b]. \tag{7.7}$$

However, both polynomials on the left side of (7.7) are positive for $x \not\varepsilon [a, b]$ so that $U_{2n}(x) = P_{2n-2}(x) + \epsilon(a - x)(b - x)Q_{n-1}^2(x)$ is a polynomial of exact degree $2n$ which is positive for $x \varepsilon (-\infty, \infty)$. We now prove that $U_{2n}(x)$ possesses a representation of the form

$$U_{2n}(x) = (\mathbf{g}(x), \mathbf{T}\mathbf{g}(x)) \tag{7.8}$$

where $\mathbf{g}(x) = (1, x, \cdots, x^n)$ and $\mathbf{T}$ is a positive definite $(n + 1) \times (n + 1)$ matrix. To this end we first write $U_{2n}(x)$ in the form $U_{2n}(x) = \sum_{k=0}^{n} V_{2k}(x)$ where $V_{2k}(x)$, $k = 0, 1, \cdots, n$, is positive on $(-\infty, \infty)$ and of exact degree $2k$, e.g., let $V_{2k}(x) = x^{2k} + \alpha$ $(\alpha > 0)$ for $k = 0, 1, \cdots, n - 1$, and $V_{2n} = U_{2n} - \delta\sum_{k=0}^{n-1} V_{2k}$ where δ is sufficiently small and positive. By Corollary VI.8.1 of Karlin and Studden (1966) each of the polynomials $V_{2k}(x)$ can be written in the form $V_{2k}(x) = R_k^2(x) + S_{k-1}^2(x)$ where R_k and S_{k-1} $(S_{-1} \equiv 0)$ are polynomials of exact degree k and $k - 1$ respectively. Now let $\mathbf{r}_{(k)}$ and $\mathbf{s}_{(k-1)}$ denote the coefficient vectors of R_k and S_{k-1}, i.e., $R_k(x) = (\mathbf{r}_{(k)}, \mathbf{g}(x))$ and $S_{k-1}(x) = (\mathbf{s}_{(k-1)}, \mathbf{g}(x))$. Since the vectors $\mathbf{r}_{(k)}$, $k = 0, 1, \cdots, n$, are linearly independent the matrix $\mathbf{T} = \sum_{k=0}^{n} [\mathbf{r}_{(k)}\mathbf{r}'_{(k)} + \mathbf{s}_{(k-1)}\mathbf{s}'_{(k-1)}]$ is positive definite and $U_{2n}(x)$ can be expressed as $U_{2n}(x) = (\mathbf{g}(x), \mathbf{T}\mathbf{g}(x))$ as required. We have thus shown that $P_{2n-2}(x) - \epsilon(x - a)(b - x)Q_{n-1}^2(x) = (\mathbf{g}(x), \mathbf{T}\mathbf{g}(x))$ where $\mathbf{T}$ is positive definite. Dividing both sides by $P_{2n-2}(x)$ we conclude that

$$(\mathbf{f}(x), \mathbf{T}\mathbf{f}(x)) = 1 - \epsilon(x - a)(b - x)Q_{n-1}^2(x)/P_{2n-2}(x)$$

which shows that $(\mathbf{f}(x), \mathbf{T}\mathbf{f}(x)) \leqq 1$ for $x \varepsilon [a, b]$ and equality holds for x in the spectrum of ξ_0. The design ξ_0 is therefore admissible by Theorem 7.1 (i).

We now assume that ξ_0 is admissible and wish to show that the spectrum of ξ_0 has at most $n - 1$ points in (a, b). Appealing to Theorem III.1.1 of Karlin and Studden (1966) we infer that if the spectrum of ξ_0 contains at least n points belonging to (a, b) then there exist a measure ξ_1, concentrating on the endpoints a and b and $n - 1$ points in (a, b), and $\lambda > 0$ such that the relations $\mu_\nu(\xi_0) = \mu_\nu(\lambda\xi_1)$, $\nu = 0, 1, \cdots, 2n - 1$, and $\mu_{2n}(\xi_0) < \mu_{2n}(\lambda\xi_1)$ hold where $\mu_\nu(\xi) = \int_a^b w(x)x^\nu\xi(dx)$, $\nu = 0, 1, \cdots, 2n$. But

$$1 = \int w(x)\sum_{i=0}^{2n-2} b_i x^i \xi_0(dx) = \sum_{i=0}^{2n-2} b_i\mu_i(\xi_0) = \sum_{i=0}^{2n-2} b_i\mu_i(\lambda\xi_1) = \lambda$$

so that $\mu_\nu(\xi_0) = \mu_\nu(\xi_1)$, $\nu = 0, \cdots, 2n-1$, and $\mu_{2n}(\xi_0) < \mu_{2n}(\xi_1)$. These relations trivially imply that ξ_0 is inadmissible. From this contradiction we conclude that the spectrum of ξ_0 consists of at most $n-1$ points in (a, b). This completes the proof.

For a general system of functions $f_0, f_1, \cdots, f_n$ the problem of characterizing the class of admissible designs seems to be quite formidable. However, for the case where $f_i(x) = [w(x)]^{\frac{1}{2}}x^i$, $i = 0, 1, \cdots, n$, $\mathfrak{X} = [a, b]$ and $w(x)$ satisfies certain conditions, an upper bound on the number of points in the spectrum of an admissible design can be obtained.

THEOREM 7.4. *Let $f_i(x) = [w(x)]^{\frac{1}{2}}x^i$, $i = 0, 1, \cdots, n$, $\mathfrak{X} = [a, b]$ and suppose $w(x)$ either satisfies one of the four conditions of Theorem 5.2 or $w(x)$ is one of the classical weight functions featured in Theorem 5.1. If a design ξ_0 is admissible then the spectrum of ξ_0 contains at most $n+1$ points.*

PROOF. If ξ_0 is admissible, then Theorem 7.1 (ii) implies the existence of a non-negative polynomial $Q_{2n}(x)$ of degree at most $2n$ such that

$$w(x)Q_{2n}(x) \leqq 1 \qquad \text{for all} \quad x \in [a, b] \tag{7.9}$$

and equality holds on the spectrum of ξ_0.

The proof proceeds by determining the number of possible zeros of the function $w(x)Q_{2n}(x) - 1$. If $w(x)$ satisfies one of the four conditions of Theorem 5.2 the analysis paraphrases that is given in the proof of Theorem 5.2 and may therefore be omitted.

Let $w(x)$ be one of the classical weight functions considered in Theorem 5.1 and suppose that the spectrum of ξ_0 contains r points. Note that any point x for which $w(x) = 0$ cannot belong to the spectrum of ξ_0. Since each of the weight functions under consideration (other than $w(x) \equiv 1$) vanishes for at least one of the endpoints a or b and $w(x)Q_{2n}(x) \not\equiv 1$ it follows from (7.9) that the function $w(x)Q_{2n}(x) - 1$ possesses at least $2r - 1$ zeros, counting multiplicities in $[a, b]$. In this circumstance the derivative

$$w'(x)Q_{2n}(x) + w(x)Q'_{2n}(x) \tag{7.10}$$

has at least $2r - 2$ zeros in the open interval (a, b). In the case of the classical weight functions the ratio $w'(x)/w(x)$ reduces to a function of the form $l(x)/q(x)$ where $l(x)$ is linear in x and $q(x)$ is a quadratic. Therefore

$$L(x) = l(x)Q_{2n}(x) + q(x)Q'_{2n}(x) \tag{7.11}$$

possesses at least $2r - 2$ zeros on (a, b). But $L(x)$ cannot vanish identically since $w(x)Q_{2n}(x) \not\equiv 1$. Therefore since $L(x)$ is a polynomial of degree at most $2n + 1$ we conclude that $2r - 2 \leqq 2n + 1$ or $r \leqq n + 1$. The proof is complete.

To testify to the difficulties involved in completely characterizing the admissible designs for a general system of functions $f_0, f_1, \cdots, f_n$ or even for the cases considered in Theorem 7.4 we cite the following examples:

(a) If $\mathfrak{X} = [-1, 1]$, $w(x) = 1 - x^2$ and $n = 1$ then a two point design concentrating at $-x_0$ and x_0 is admissible if $0 < x_0 < 1/2^{\frac{1}{2}}$ and inadmissible when $1/2^{\frac{1}{2}} < x_0 \leqq 1$.

(b) For the case $w(x) = e^{-x}$ and $\mathfrak{X} = [0, \infty]$ it can be shown that at most n points in the open interval $(0, \infty)$ are permitted in the spectrum of any admissible design. For the special case $n = 1$ a design concentrating on the points $x_0 = 0$ and x_1 is admissible provided $x_1 < \gamma$ where γ is the unique positive root of the equation $e^{\gamma}(\gamma - 2)^2 = 4$ and inadmissible when $x_1 > \gamma$.

8. Quadratic loss. In this section we will be concerned with the minimization of the expectation $\mathcal{E}\mathbf{W}$ where $\mathbf{W} = ((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}))$ and $\mathbf{L}$ is a non-negative matrix of rank k. The estimates $\hat{\theta}_i$, $i = 0, 1, \cdots, n$, will be assumed to be linear in the observations. Let $\xi = \{x_i ; p_i\}_1^r$ where $p_i N = n_i$ are integers and let $\mathbf{A}$ denote the $N \times (n + 1)$ matrix with n_i rows equal to $\mathbf{f}(x_i) = (f_0(x_i), \cdots, f_n(x_i))$, $i = 1, \cdots, r$. If $\hat{\boldsymbol{\theta}} = \mathbf{By}$, where $\mathbf{y}$ is the vector of N observations and $\mathbf{B}$ is an arbitrary $(n + 1) \times N$ matrix, then

$$(8.1) \qquad \mathcal{E}\mathbf{W} = \boldsymbol{\theta}'(\mathbf{BA} - \mathbf{I})'\mathbf{L}(\mathbf{BA} - \mathbf{I})\boldsymbol{\theta} + \operatorname{tr} \mathbf{B}'\mathbf{LB}.$$

In order to make the estimates $\hat{\boldsymbol{\theta}}$ unbiased in the sense that the right side of (8.1) is independent of $\boldsymbol{\theta}$ we restrict considerations to those designs ξ whose associated matrix $\mathbf{A}$ is such that there exists a matrix $\mathbf{B}$ satisfying

$$(\mathbf{BA} - \mathbf{I})'\mathbf{L}(\mathbf{BA} - \mathbf{I}) = \mathbf{0}.$$

Since $\mathbf{L}$ is of rank k we write $\mathbf{L} = \sum_{\nu=1}^{k} \mathbf{l}_\nu \mathbf{l}_\nu'$ where $\mathbf{l}_1, \cdots, \mathbf{l}_k$ are linearly independent vectors and let $\mathbf{J}$ denote the $n + 1 \times k$ matrix with column vectors $\mathbf{l}_\nu$, $\nu = 1, \cdots, k$. Then $\mathbf{JJ}' = \mathbf{L}$ and the restriction that there exist a solution of $(\mathbf{BA} - \mathbf{I})'\mathbf{L}(\mathbf{BA} - \mathbf{I}) = 0$ is equivalent to the existence of a solution $\mathbf{B}$ to the matrix equation $\mathbf{J}'\mathbf{BA} = \mathbf{J}'$. Let $\mathcal{B}$ denote the class of $n + 1 \times k$ matrices $\mathbf{B}$ satisfying $\mathbf{J}'\mathbf{BA} = \mathbf{J}'$. For each $\mathbf{B} \in \mathcal{B}$, Equation (8.1) reduces to

$$(8.2) \qquad \mathcal{E}\mathbf{W} = \operatorname{tr} \mathbf{B}'\mathbf{LB}.$$

We now minimize (8.2) over the class $\mathcal{B}$. Let $\mathbf{E}$ be an arbitrary $n + 1 \times k$ matrix and assume that $\mathbf{B}$ belongs to $\mathcal{B}$. With the aid of Schwartz's inequality we obtain

$$(8.3) \qquad \begin{aligned} \operatorname{tr}^2 \mathbf{E}'\mathbf{J} = \operatorname{tr}^2 \mathbf{E}'\mathbf{A}'\mathbf{B}'\mathbf{J} &\leqq (\operatorname{tr} \mathbf{E}'\mathbf{A}'\mathbf{AE})(\operatorname{tr} \mathbf{B}'\mathbf{JJ}'\mathbf{B}) \\ &= N(\operatorname{tr} \mathbf{E}'\mathbf{M}(\xi)\mathbf{E})(\operatorname{tr} \mathbf{B}'\mathbf{LB}). \end{aligned}$$

Moreover equality occurs if and only if $\mathbf{AE}$ is proportional to $\mathbf{B}'\mathbf{J}$.

Let $\lambda_1, \cdots, \lambda_s$ be the non-zero eigenvalues of $N^{-1}\mathbf{A}'\mathbf{A} = \mathbf{M}(\xi)$ and $\varphi_1, \cdots, \varphi_s$ the associated orthonormal eigenvectors. If

$$(8.4) \qquad \mathbf{E}_0 = N^{-1}\left(\textstyle\sum_{i=1}^{s} \lambda_i^{-1}\varphi_i\varphi_i'\right)\mathbf{J} \quad \text{and} \quad \mathbf{B}_0 = N^{-1}\left(\textstyle\sum_{i=1}^{s} \lambda_i^{-1}\varphi_i\varphi_i'\right)\mathbf{A}'$$

then clearly $\mathbf{AE}_0 = \mathbf{B}_0'\mathbf{J}$ so that equality applies in (8.3) when $\mathbf{E}_0$ and $\mathbf{B}_0$ are specified as in (8.4).

Now if $U = \{\mathbf{e} \mid \mathbf{M}(\xi)\mathbf{e} = \mathbf{0}\}$ it follows from (8.3) that the columns of $\mathbf{J}$ belong to $U^{\perp}$ so that $(\sum_{i=1}^{s} \varphi_i\varphi_i')\mathbf{J} = \mathbf{J}$. In this case $\mathbf{A}'\mathbf{B}_0'\mathbf{J} = N^{-1}(\mathbf{A}'\mathbf{A}) \cdot (\sum_{i=1}^{s} \lambda_i^{-1}\varphi_i\varphi_i')\mathbf{J} = (\sum_{i=1}^{s} \varphi_i\varphi_i')\mathbf{J} = \mathbf{J}$ and hence $\mathbf{B}_0 \in \mathcal{B}$. From (8.3) we then deduce that

$$\inf_{\mathbf{B}\varepsilon\mathcal{B}} \operatorname{tr} \mathbf{B}'\mathbf{L}\mathbf{B} = N^{-1} \sup_{\mathbf{E}} [\operatorname{tr}^2 \mathbf{E}'\mathbf{J}/\operatorname{tr} \mathbf{E}'\mathbf{M}(\xi)\mathbf{E}] \tag{8.5}$$

where the ratio in the right hand term is defined to be zero in the event that $\operatorname{tr} \mathbf{E}'\mathbf{M}(\xi)\mathbf{E} = \mathbf{0}$.

We have thus proven:

THEOREM 8.1. *Let* $\mathbf{L} = \sum_{\nu=1}^{k} \mathbf{l}_\nu \mathbf{l}_\nu'$, *where* $\mathbf{l}_1, \cdots, \mathbf{l}_k$ *are linearly independent, denote a positive semi-definite matrix of rank* k *and let* $\xi = \{x_i ; p_i\}_1^r$ *where* $p_i N = n_i$ *are integers. Let* $\mathbf{W} = ((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \mathbf{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}))$ *and define* $\mathbf{A}$, $\mathbf{J}$ *and* $\mathcal{B}$ *as above. If* $\mathcal{B}$ *is non-empty then*

$$\inf_{\mathbf{B}\varepsilon\mathcal{B}} \mathcal{E}\mathbf{W} = \inf_{\mathbf{B}\varepsilon\mathcal{B}} \operatorname{tr} \mathbf{B}'\mathbf{L}\mathbf{B} = N^{-1} \sup_{\mathbf{E}} [\operatorname{tr}^2 \mathbf{E}'\mathbf{J}/\operatorname{tr} \mathbf{E}'\mathbf{M}(\xi)\mathbf{E}]. \tag{8.6}$$

REMARK 8.1. (i) If the matrix $\mathbf{L}$ has rank one and $\mathbf{L} = \mathbf{l}\mathbf{l}'$ then (8.6) reduces to

$$\inf \mathcal{E}\mathbf{W} = \sup_{\mathbf{e}} [(\mathbf{e}, \mathbf{l})^2/N(\mathbf{e}, \mathbf{M}(\xi)\mathbf{e})]. \tag{8.7}$$

This is precisely the expression for the minimum variance among all linear unbiased estimates of $(\mathbf{l}, \boldsymbol{\theta})$ using the design ξ (see Theorem 2.1).

(ii) Using the value for either $\mathbf{B}_0$ or $\mathbf{E}_0$ from (8.4) the expression (8.6) may also be put in the form

$$N^{-1} \sup_{\mathbf{E}} [\operatorname{tr}^2 \mathbf{E}'\mathbf{J}/\operatorname{tr} \mathbf{E}'\mathbf{M}(\xi)\mathbf{E}] = N^{-1} \sum_{i=1}^{k} \sum_{j=1}^{s} \lambda_j^{-1} (\varphi_j, \mathbf{l}_i)^2.$$

Note that this is the sum of the minimum variances for estimating $(\mathbf{l}_i, \boldsymbol{\theta})$ using the design ξ (see Theorem 2.1).

(iii) In the case where $\mathbf{M}(\xi)) > \mathbf{0}$ Equation (8.6) becomes $N^{-1} \operatorname{tr} \mathbf{L}\mathbf{M}^{-1}(\xi)$ and $\mathbf{E}_0 = N^{-1}\mathbf{M}^{-1}(\xi)\mathbf{J}$ and $\mathbf{B}_0 = N^{-1}\mathbf{M}^{-1}(\xi)\mathbf{A}'$.

Our next objective is to give a partial characterization of the designs ξ which minimize the expression

$$\sup_{\mathbf{E}} [\operatorname{tr}^2 \mathbf{E}'\mathbf{J}/\operatorname{tr} \mathbf{E}'\mathbf{M}(\xi)\mathbf{E}] \tag{8.8}$$

over the class of all probability measures ξ. The supremum in the above expression is evaluated with respect to the set of all non-zero matrices $\mathbf{E}$ whose column vectors belong to $U^\perp$ where $U = U(\xi) = \{\mathbf{e} \mid \mathbf{M}(\xi)\mathbf{e} = \mathbf{0}\}$. Note that (8.8) is finite if and only if the column vectors of the matrix $\mathbf{J}$ lie in $U^\perp$. In the following we restrict attention exclusively to the set of designs ξ for which (8.8) is finite.

Clearly (8.8) is minimized for an admissible design ξ_0. Therefore by Theorem 7.2(ii) there exists a non-negative matrix $\mathbf{T}$ such that $(\mathbf{f}(x), \mathbf{T}\mathbf{f}(x)) \leqq 1$ for all $x \varepsilon \mathfrak{X}$ and equality holds for x in the spectrum of ξ_0. It will be demonstrated that the matrix $\mathbf{T}$ may be determined to have the same rank k as the loss matrix $\mathbf{L}$. The following lemma is needed.

LEMMA 8.1. *Let* $\mathbf{J}$ *denote an arbitrary* $n + 1 \times k$ *matrix of rank* k *with column vectors* $\mathbf{l}_\nu$, $\nu = 1, \cdots, k$. *For an arbitrary design* ξ *let* $\lambda_1, \cdots, \lambda_s$ *denote the non-zero eigenvalues of* $\mathbf{M}(\xi)$ *with associated orthornormal eigenvectors* $\varphi_1, \cdots, \varphi_s$.

Then

$$\sup_{\mathbf{E}} [\mathrm{tr}^2\, \mathbf{E}'\mathbf{J}/\mathrm{tr}\, \mathbf{E}'\mathbf{M}(\xi)\mathbf{E}] = \sum_{i=1}^{k} \sum_{j=1}^{s} \lambda_j^{-1}(\varphi_j, \mathbf{l}_i)^2 \tag{8.9}$$

and equality is achieved if and only if the columns of $\mathbf{E}$ *are proportional to*

$$\mathbf{e}_i^0 = \sum_{j=1}^{s} \lambda_j^{-1}(\varphi_j, \mathbf{l}_i)\varphi_j, \qquad i = 1, 2, \cdots, k. \tag{8.10}$$

Proof. Let $\mathbf{e}_i$ belong to $U^{\perp}$ where $U = \{\mathbf{e} \mid \mathbf{M}(\xi)\mathbf{e} = \mathbf{0}\}$. Applying Schwartz's inequality twice we obtain (see 2.7)

$$(\mathbf{e}_i, \mathbf{l}_i) \leqq [(\mathbf{e}_i, \mathbf{M}(\xi_0)\mathbf{e}_i)(\sum_{j=1}^{s} \lambda_j^{-1}(\varphi_j, \mathbf{l}_i)^2)]^{\frac{1}{2}} \tag{8.11}$$

and

$$[\sum_{i=1}^{k} (\mathbf{l}_i, \mathbf{e}_i)]^2 \leqq \sum_{i=1}^{k} (\mathbf{e}_i, \mathbf{M}(\xi_0)\mathbf{e}_i) \sum_{i=1}^{k} \sum_{j=1}^{s} \lambda_j^{-1}(\varphi_j, \mathbf{l}_i)^2. \tag{8.12}$$

Moreover, equality occurs in (8.11) and also in (8.12) if and only if $\mathbf{e}_i$, $i = 1, \cdots, k$, is proportional to

$$\mathbf{e}_i^0 = \sum_{j=1}^{s} \lambda_j^{-1}(\varphi_j, \mathbf{l}_i)\varphi_j \quad , \qquad i = 1, 2, \cdots, k. \tag{8.13}$$

In this case

$$\sup_{\mathbf{E}} [\mathrm{tr}^2\, \mathbf{E}'\mathbf{J}/\mathrm{tr}\, \mathbf{E}'\mathbf{M}(\xi_0)\mathbf{E}] = \sum_{i=1}^{k} \sum_{j=1}^{s} \lambda_j^{-1}(\varphi_j, \mathbf{l}_i)^2$$

and equality is achieved only when the columns of $\mathbf{E}$ are given by (8.13).

The following theorem is stated in Elfving (1959) without proof. The proof indicated below emphasizes game theoretic arguments in the spirit of the paper hereto.

Theorem 8.2. *Let* $\mathbf{J}$ *denote an* $n + 1 \times k$ *matrix of rank* k. *If* ξ_0 *satisfies*

$$v = \inf_{\xi} \sup_{\mathbf{E}} [\mathrm{tr}^2\, \mathbf{E}'\mathbf{J}/\mathrm{tr}\, \mathbf{E}'\mathbf{M}(\xi)\mathbf{E}] = \sup_{\mathbf{E}} [\mathrm{tr}^2\, \mathbf{E}'\mathbf{J}/\mathrm{tr}\, \mathbf{E}'\mathbf{M}(\xi_0)\mathbf{E}] \tag{8.14}$$

then (i) *there exists a non-negative matrix* $\mathbf{T}$ *of rank* k *such that* $(\mathbf{f}(x), \mathbf{Tf}(x)) \leqq 1$ *for all* $x \in \mathfrak{X}$ *with equality occurring for* x *in the spectrum of* ξ_0;

(ii) *when* $|\mathbf{M}(\xi_0)| > 0$ *the matrix* $\mathbf{T}$ *may be chosen to be* $\mathbf{T} = h^{-2}\mathbf{M}^{-1}(\xi_0)\mathbf{L}\mathbf{M}^{-1}(\xi_0)$ *where* $h^2 = \mathrm{tr}\, \mathbf{L}\mathbf{M}^{-1}(\xi_0)$.

Proof. Consider the sequence of games with kernels

$$\phi_\epsilon(\mu, \xi) = \int [\mathrm{tr}^2\, \mathbf{E}'\mathbf{J}/\mathrm{tr}\, \mathbf{E}'\mathbf{M}(\xi)\mathbf{E}]\, d\mu(\mathbf{e}_1, \cdots, \mathbf{e}_k)$$

where $\mathbf{E}$ is an arbitrary $n + 1 \times k$ matrix with column vectors $\mathbf{e}_1, \cdots, \mathbf{e}_k$ respectively such that $\sum \|\mathbf{e}_i\|^2 = 1$, μ is a probability measure defined on the above set of matrices $\mathbf{E}$ and ξ is such that $\mathbf{M}(\xi)$ has eigenvalues all of which are not less than ϵ.

Each of these games determines a value $v_\epsilon \geqq v$ (see (8.14)) and possess optimal strategies, say $\mathbf{M}(\xi_\epsilon)$ and μ_ϵ. Moreover μ_ϵ must concentrate on the single matrix $\mathbf{E}_\epsilon$ which is proportional to $\mathbf{M}^{-1}(\xi_\epsilon)\mathbf{J}$.

Choose a sequence $\epsilon_n \to 0$ so that $\mathbf{E}_{\epsilon_n} \to \mathbf{E}_*$ and $v_{\epsilon_n} \to v_0$. Using the fact that $v_0 \geqq v$ we deduce that

$$\mathrm{tr}^2\, \mathbf{E}_*\mathbf{J}' \geqq v\, \mathrm{tr}\, \mathbf{E}_*'\mathbf{M}(\xi)\mathbf{E}_* \qquad \text{for all} \quad \xi. \tag{8.15}$$

In particular for ξ concentrating on a single point $x \varepsilon \mathfrak{X}$ we obtain

$$(8.16) \qquad \mathrm{tr}^2\, \mathbf{E}_*\mathbf{J}' \geqq v(\mathbf{f}(x), \mathbf{E}_*\mathbf{E}_*'\mathbf{f}(x)) \qquad \text{for all} \quad x \,\varepsilon\, \mathfrak{X}.$$

Now from (8.14) we infer that

$$(8.17) \qquad \mathrm{tr}^2\, \mathbf{E}_*'\mathbf{J} \leqq v \,\mathrm{tr}\, \mathbf{E}_*'\mathbf{M}(\xi_0)\mathbf{E}_*\,.$$

Comparing (8.15) and (8.17) we see that equality necessarily holds in (8.15) when $\xi = \xi_0$ and hence equality occurs in (8.16) for x in the spectrum of ξ_0.

Since $v > 0$ and $\mathrm{tr}\, \mathbf{E}_*'\mathbf{M}(\xi)\mathbf{E}_* > 0$ for $|\mathbf{M}(\xi)| > 0$ it follows from (8.15) that $\mathrm{tr}^2\, \mathbf{E}_*'\mathbf{J} > 0$. We may therefore define $h^2 = v^{-1}\,\mathrm{tr}^2\, \mathbf{E}_*\mathbf{J}$ and set $\mathbf{T} = h^{-2}\mathbf{E}_*\mathbf{E}_*'$. The assertion of part (i) then follows with this choice of $\mathbf{T}$ provided $\mathbf{E}_*$ is of rank k. But

$$\mathrm{tr}^2\, \mathbf{E}_*'\mathbf{J}/\mathrm{tr}\, \mathbf{E}_*'\mathbf{M}(\xi_0)\mathbf{E}_* = \sup_{\mathbf{E}}\, [\mathrm{tr}^2\, \mathbf{E}\mathbf{J}/\mathrm{tr}\, \mathbf{E}'\mathbf{M}(\xi_0)\mathbf{E}\,]$$

and hence by Lemma 8.1 the columns of $\mathbf{E}_*$ must be proportional to the vectors (8.13) associated with ξ_0. These vectors are readily discerned to be linearly independent since the vectors $\mathbf{l}_\nu$, $\nu = 1, 2, \cdots, k$, were assumed to be linearly independent.

It remains to verify part (ii). However when $|\mathbf{M}(\xi_0)| > 0$ the matrix $\mathbf{E}_*$ is proportional to $\mathbf{M}^{-1}(\xi_0)\mathbf{J}$ so that we may take

$$\mathbf{T} = h^{-2}\mathbf{E}_*\mathbf{E}_*' = h^{-2}\mathbf{M}^{-1}(\xi_0)\mathbf{L}\mathbf{M}^{-1}(\xi_0)$$

where $h^2 = v^{-1}\,\mathrm{tr}^2\, \mathbf{E}_*\mathbf{J}' = \mathrm{tr}\, \mathbf{L}\mathbf{M}^{-1}(\xi_0)$.

REFERENCES

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 255–262.

ELFVING, G. (1959). *Design of Linear Experiments, Probability and Statistics* (Cramér volume) 58–74 Wiley, New York.

HOEL, P. G., and LEVINE, A. (1964). Optimal spacing and weighing in polynomial regression. *Ann. Math. Statist.* **35** 1553–1560.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics*. Interscience, New York.

KIEFER, J. (1959). Optimum experimental designs, *J. Roy. Statist. Soc. Ser. B* **21** 273–319.

KIEFER, J. (1960). Optimum experimental designs V. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 381–405. Univ. of California Press.

KIEFER, J. (1962). An extremum result. *Can. J. Math.* **14** 597–601.

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

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

KIEFER, J. and WOLFOWITZ, J. (1965). On a theorem of Hoel and Levine on extrapolation designs. *Ann. Math Statist.* **36** 1627–1655

SCHOENBERG, I. J. (1951). On Pólya frequency functions. *J. d'Analyze Math.* **1** 331–374.

SCHOENBERG, I. J. (1959). On the maxima of certain Hankel determinants and the zeros of the classical orthogonal polynomials. *Nederl. Akad. Netensch. Proc. Ser.* A62-*Indag. Math.* **21** 282–290.

SZEGÖ, G. (1959). *Orthogonal Polynomials* (revised edition) Amer. Math. Soc. Colloq. Publ. **23** New York.