statistics

MATHEMATISCHE OPERATIONSFORSCHUNG UND STATISTIK

SERIES STATISTICS

Volume 12
1981
Number 1–4

January/December

Published at
Institute of Mathematics
Academy of Sciences of the G.D.R.

AKADEMIE-VERLAG · BERLIN

Math. Operationsforsch. Statist., Ser. Statistics, Vol. 12 (1981) No. 1, pp. 13–20

On c-Optimal Design Measures

FRIEDRICH PUKELSHEIM [1]

Summary. A short proof is presented to construct c-optimal measures for designing experiments, and emphasis is laid on the geometry inherent in this problem. A number of examples and counterexamples are given which relate to other results in the literature.

Key words: Approximate design theory, ELFVING's Theorem, Singular information matrices.

1. Introduction and notation

The paper characterizes design measures ξ which are in a certain sense optimal for estimating a linear form $c'\beta$, or testing a linear hypothesis $c'\beta=0$, where the k parameters $\beta_1, \ldots, \beta_k$ form the unknown vector β, and c is some prescribed vector in R^k. With the usual further assumptions as detailed below, the present approach draws heavily on convex geometry in order to arrive at the desired conclusions. In Section 2 (*Elfving's Theorem*) the classical result of ELFVING (1952) is rederived in a way that shortens and clarifies the argument of SIBSON (1974). In Section 3 (*Elfving Charts*) the transparency thus gained leads to some examples pointing to possible extensions, and some counterexamples refuting a number of assertions that can be found in the literature.

In the theory of experimental design one has in mind that the experimenter can choose a *level* x in a *design space* $\mathfrak{X}$, and then make uncorrelated observations having equal variance independent of x and expectation $f(x)'\beta$ where the *regression function* f is supposed to be known and fixed. A *design* ξ specifies the levels x to be chosen and the proportions $\xi(x)$ of observations to be drawn at x. A c-optimal design then ensures smallest possible variance of the least squares estimate for $c'\beta$, or maximal power of the F-test for $c'\beta=0$. The textbooks FEDOROV (1972), BANDEMER u. a. (1977), HUMAK (1977), or KRAFFT (1978) present the full background and extensive bibliographies on optimal design of experiments. Of the underlying assumptions we here give only such details as are necessary.

The *design space* $\mathfrak{X}$ is supposed to be a measurable space with measurable one-point sets $\{x\}$, and Ξ is the set of all probability measures, called *design*

[1] Inst. Math. Stochast. Univ., Hermann-Herder-Str. 10, D-7800 Freiburg i. B., Fed. Rep. of Germany.

measures, ξ on $\mathfrak{X}$. The *regression function* f is assumed to be a measurable function from $\mathfrak{X}$ into R^k such that its image $f(\mathfrak{X})$ is compact and spans all of R^k. The *information matrix* of a design measure ξ is defined by

$$\operatorname{Info}(\xi) = \int_{\mathfrak{X}} f(x)\, f(x)'\, d\xi ,$$

and the set of all these matrices is denoted by Info (Ξ). For a given R^k-vector c considerations of estimability of $c'\beta$, or testability of $c'\beta=0$, direct interest towards the set

$$\mathfrak{A}(c) = \{A \in NND(k) \mid c \in \text{range } A\} ,$$

i.e., the set of all real symmetric non-negative definite $k \times k$ matrices A such that $c = Az$ for some $z \in \mathrm{R}^k$. Then a design measure ξ is called *optimal for* $c'\beta$ if its information matrix M lies in $\mathfrak{A}(c)$ and minimizes the real function $\Phi(M) = c'M^-c$. As usual, a prime stands for transposition, and a g-inverse A^- means any matrix satisfying $M(M^-)M = M$. On $\mathfrak{A}(c)$ the function Φ is well defined and positive.

2. ELFVING's theorem

First construct the symmetric convex set

$$\mathfrak{R} = \text{convex hull of } f(\mathfrak{X}) \cup -f(\mathfrak{X}) .$$

In fact, $\mathfrak{R}$ is also compact (ROCKAFELLAR, 1970, Theorem 17.2, p. 158), and since $f(\mathfrak{X})$ spans R^k the interior of $\mathfrak{R}$ is non-empty and contains 0. Secondly, associate with every vector $c \in \mathrm{R}^k$ the smallest number $\mu \geqq 0$ such that c is contained in the *regression ball* $\mu\mathfrak{R} = \{\mu v \mid v \in \mathfrak{R}\}$, i.e.,

$$\varrho(c) = \inf\{\mu \geqq 0 \mid c \in \mu\mathfrak{R}\} .$$

Thus ϱ is a norm, and $\mathfrak{R}$ coincides with its unit ball $\{v \in \mathrm{R}^k \mid \varrho(v) \leqq 1\}$ (ROCKAFELLAR, 1970, Theorem 15.2, p. 131). Because of their close relation to the regression function f we shall call $\mathfrak{R}$ the *regression ball* and ϱ the *regression norm.*

As a member of the regression ball $\mathfrak{R}$ the vector $c/\varrho(c)$ is a convex combination of finitely many points in $f(\mathfrak{X})$ or $-f(\mathfrak{X})$: for every $c \in \mathrm{R}^k$ there exists a natural number n, and points $x_i \in \mathfrak{X}$, $\varepsilon_i \in \{\pm 1\}$, and $\lambda_i > 0$ $(i = 1, \ldots, n)$ such that $\sum \lambda_i = 1$ and

$$c = \varrho(c) \sum_{i=1}^{n} \lambda_i \varepsilon_i f(x_i). \tag{2.1}$$

Although derived from the geometric point of view, representation (2.1) also contains the solution to the optimal design problem. This will be made precise in Theorem 1 which dates back to ELFVING (1952); see also KARLIN & STUDDEN (1966, p. 789). The proof below follows SILVEY's idea (WYNN, 1972, p. 174; SIBSON, 1974) the set of all cylinders, including ellipsoids, containing the regression ball $\mathfrak{R}$, namely,

$$\operatorname{cyl}(\mathfrak{R}) = \{N \in NND(k) \mid v'Nv \leqq 1 \quad \text{for all} \quad v \in \mathfrak{R}\} ,$$

rather than the quite arbitrary convex body $\mathfrak{R}$ alone.

Theorem 1. *Given $c \in \mathrm{R}^k$, the infimum of $c'M^-c$ among all $M \in \mathfrak{A}(c) \cap \mathrm{Info}\ (\Xi)$ is equal to $\{\varrho(c)\}^2$ and is attained. In fact, whenever c is represented as in (2.1) then the design measure ξ defined by $\xi(x_i)=\lambda_i$ $(i=1, \ldots, n)$ is optimal for $c'\beta$.*

Proof. The proof is in two steps. First it is shown that $c'M^-c \geqq c'Nc$ holds for two arbitrary members $M=\mathrm{Info}\ (\xi) \in \mathfrak{A}(c) \cap \mathrm{Info}\ (\Xi)$, and $N \in \mathrm{cyl}\ (\mathfrak{R})$. For when $1 \geqq f(x)'Nf(x)$ is integrated with respect to ξ then one arrives at $1 \geqq \mathrm{trace}\ MN = \|M^{\frac{1}{2}}N^{\frac{1}{2}}\|^2$. Here $\|.\|$ is the norm associated with the Euclidean matrix inner product $\langle A, B\rangle = \mathrm{trace}\ A'B$ on the space $\mathrm{R}^{k\times k}$ of all real $k\times k$ matrices. An orthogonal projection is obtained when A is mapped into $M^{\frac{1}{2}+}c(M^{\frac{1}{2}+}c)^+A$, and therefore $\|M^{\frac{1}{2}}N^{\frac{1}{2}}\|^2 \geqq \|M^{\frac{1}{2}+}c(M^{\frac{1}{2}+}c)^+M^{\frac{1}{2}}N^{\frac{1}{2}}\|^2$. Since $M \in \mathfrak{A}(c)$, the property $A^+ = (A'A)^+A'$ of a Moore-Penrose inverse A^+ proves the last term to be $c'Nc(c'M^-c)^{-1}$, as desired.

The second step follows Sibson (1974, p. 691) and shows that when M is the information matrix $\Sigma\ \lambda_i f(x_i)\ f(x_i)'$ of the measure ξ defined in the theorem then there exists some $N \in \mathrm{cyl}\ (\mathfrak{R})$ with $c'M^-c=\{\varrho(c)\}^2=c'Nc$, thus proving the assertions. For let $d \in \mathrm{R}^k$ define the hyperplane $\{u \in \mathrm{R}^k \mid d'u=1\}$ supporting $\mathfrak{R}$ in its boundary point $c/\varrho(c)$, i.e., $d'v \leqq 1 = d'c/\varrho(c)$ for all $v \in \mathfrak{R}$. From $\varepsilon_i d'f(x_i) \leqq 1$ and (2.1) one gets $\varepsilon_i = d'f(x_i)$. With Whittle's (1973, p. 129) quasilinear representation one then has the following:

$$\begin{aligned} c'M^-c &= \sup_{h\in \mathrm{R}^k}\ 2h'c - h'Mh \\ &= \sup_{h\in \mathrm{R}^k}\ 2\varrho(c)\ \Sigma\ \lambda_i\varepsilon_i h'f(x_i) - \Sigma\ \lambda_i\{h'f(x_i)\}^2 \\ &= \sup_{h\in \mathrm{R}^k}\ \Sigma\ \lambda_i\ [\{\varrho(c)\ \varepsilon_i\}^2 - \{\varrho(c)\ \varepsilon_i - h'f(x_i)\}^2] \\ &= \{\varrho(c)\}^2 - \inf_{h\in \mathrm{R}k}\ \Sigma\ \lambda_i\{(\varrho(c)\ d-h)'f(x_i)\}^2. \end{aligned}$$

The infimum 0 is attained at $h=\varrho(c)\ d$. On the other hand $N=dd' \in \mathrm{cyl}\ (\mathfrak{R})$, and $c'Nc=(c'd)^2=\{\varrho(c)\}^2$. ■

Theorem 1 does not extend to linear functions $K'\beta$, with a $k\times s$ matrix K of rank $s>1$. For K, unlike c, does not admit a natural embedding in the space R^k where f takes its values. The formulation that carries over to greater generality is the following, cf., Pukelsheim (1980, Theorems 3, 4).

Corollary 2. *Given $c \in \mathrm{R}^k$, a design measure ξ with information matrix M is optimal for $c'\beta$ if and only if M lies in $\mathfrak{A}(c)$ and there exists some N in $\mathrm{cyl}\ (\mathfrak{R})$ such that $c'M^-c=c'Nc$. This equality occurs if and only if $f(x)'Nf(x)=1$ ξ-almost surely and*

$$MN = c(c'M^-c)^{-1}c'N\ . \tag{2.2}$$

Proof. See the proof of Theorem 1. The second part follows from examining equality in $1 \geqq \operatorname{trace} MN \geqq c'Nc(c'M^{-}c)^{-1}$. ■

Bounds for the regression norm ϱ can be established from the radii of the Euclidean balls inscribed in and circumscribing the regression ball $\mathfrak{R}$, whose lengths are $r = \min\{\|v\| \mid \varrho(v)=1\}$ and $R = \max\{\|v\| \mid \varrho(v)=1\}$, respectively. It is easy to see that

$$\|c\|/R \leqq \varrho(c) \leqq \|c\|/r . \tag{2.3}$$

In fact, as c varies over the Euclidean unit sphere in R^k the values $\varrho(c)$ attain every number between R^{-1} and r^{-1}. The number R also equals $\max\{\|f(x)\| \mid x \in \mathfrak{X}\}$, and this is in many cases easy to compute. Corollary 3 describes what happens when the upper bound in (2.3) is attained.

Corollary 3. *Let $c \in \mathrm{R}^k$ be an in-ball radius of $\mathfrak{R}$, i.e., $\varrho(c)=1$ and $\|c\|=r$. Then c is an eigenvector associated with eigenvalue r^2 of every information matrix M that belongs to an optimal design measure for $c'\beta$.*

Proof. Because c is an in-ball radius the hyperplane $\{u \in \mathrm{R}^k \mid c'u/r^2 = 1\}$ supports $\mathfrak{R}$ in c. By the proof of Theorem 1 equation (2.2) is satisfied for $N = cc'/r^4$, leading to $Mc = r^2 c$. ■

3. Elfving charts

Example 1 was recently proposed by SILVEY (1978, p. 554). Another example of the same type was put forward by KIEFER (1961, p. 309).

Example 1. $\mathfrak{X} = \left\{\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}4\\1\end{pmatrix}, \begin{pmatrix}4\\2\end{pmatrix}\right\}$, $f(x)=x$, $c=\begin{pmatrix}1\\0\end{pmatrix}$; see Fig. 1. The regression ball $\mathfrak{R}$ is the quadrangle with vertices $\pm\begin{pmatrix}4\\1\end{pmatrix}$ and $\pm\begin{pmatrix}4\\2\end{pmatrix}$. The prolongation of c that meets the boundary of $\mathfrak{R}$ is $c/\varrho(c) = (2/3)\begin{pmatrix}4\\1\end{pmatrix} - (1/3)\begin{pmatrix}4\\2\end{pmatrix}$, whence $\varrho(c)=3/4$. Therefore $\xi\begin{pmatrix}4\\1\end{pmatrix} = 2/3$, $\xi\begin{pmatrix}4\\2\end{pmatrix} = 1/3$ is optimal for $c'\beta$, its information matrix M is non-singular, and $\{\varrho(c)\}^2 = 9/16$.

It is now easy to give a correct formulation of Korollar 8.3.1 in HUMAK (1977, p. 453): Given $x_0 \in \mathfrak{X}$, the one-point measure $\xi(x_0)=1$ is optimal for $f(x_0)'\beta$ if and only if $f(x_0)$ lies on the boundary of the regression ball $\mathfrak{R}$. Observe that when $c/\varrho(c)$ is an extreme point of $\mathfrak{R}$ then it has a representation $\varepsilon f(x_0)$ for some $x_0 \in \mathfrak{X}$ and $\varepsilon \in \{\pm 1\}$.

Continuing the discussion of Example 1 note that SILVEY (op. cit.) showed that the one-point measure $\xi_1(c)=1$ is sub-optimal, in the terminology of HUMAK (1977,

p. 436). The range of its information matrix M_1 is certainly contained in the range R^2 of the non-singular matrix M. But at the same time $c'M_1 c = 1 > c'M_1^- c = 9/16$, contrary to Satz 8.2.3 in HUMAK (1977, p. 442).

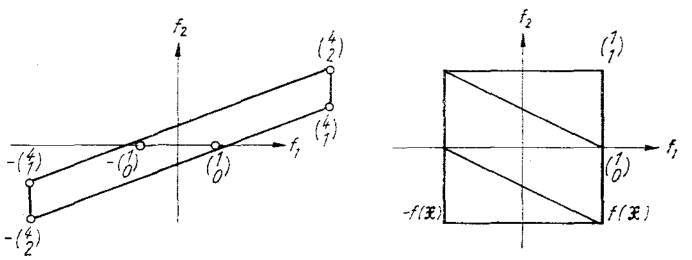

Fig. 1. The left chart refers to Example 1: the regression ball $\mathfrak{R}$ is a parallelogram. Although $c = (1, 0)'$ lies in $\mathfrak{X}$ it differs from the point where the half-ray $\{\alpha c \mid \alpha > 0\}$ intersects the boundary of $\mathfrak{R}$, namely $c/\varrho(c) = (2/3)\ (4, 1)' - (1/3)\ (4, 2)'$. It is the latter representation which proves the design $\xi((4, 1)') = 2/3$, $\xi((4, 2)') = 1/3$ to be optimal for $c'\beta$.

The right chart refers to Example 2: $\mathfrak{R}$ is a square. Observe that $c = (1, 1)'$ is an extreme point of $\mathfrak{R}$. In case $c = (1, 0)'$ a design ξ is optimal for $c'\beta$ whenever $\int x \, d\xi = 0$. The parallelogram inscribed in $\mathfrak{R}$ refers to Example 3.

Example 1 also serves as a counterexample for the equivalence theorems of Section 5.6.3 in BANDEMER u. a. (1977, pp. 220–223). For although ξ_1 is not optimal for $c'\beta$ one has $\sup f(x)'M_1^- c(c'M_1^- c)^{-1} c'M_1^- f(x) = 1$, when the supremum is taken over all $\{x \in \mathfrak{X} \mid f(x) \in \text{range } M_1\} = \left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\right\}$. ■

FEDOROV & MALYUTOV (1972, p. 286) employed the Moore-Penrose inverse M^+ and conjectured that optimality holds if and only if the supremum of $f(x)'M^+c(c'M^-c)^{-1}c'M^+f(x)$, taken over all $x \in \mathfrak{X}$, is 1. This is not so, in general, as we illustrate next.

Example 2. $\mathfrak{X} = [-1, +1]$, $f(x) = \begin{pmatrix}1\\x\end{pmatrix}$, $c = \begin{pmatrix}1\\\frac{1}{2}\end{pmatrix}$; see Fig. 1. An optimal design measure for $c'\beta$ is $\xi\begin{pmatrix}1\\\frac{1}{2}\end{pmatrix} = 1$, with optimal value $\{\varrho(c)\}^2 = 1$. However, for its information matrix M one gets $f(1)'\, M^+c(c'M^-c)^{-1}\, c'M^+f(1) = 36/25 > 1$. Note that this situation is different when $c = \begin{pmatrix}1\\1\end{pmatrix}$, as in BANDEMER u. a. (1977, p. 217), or when $c = \begin{pmatrix}1\\0\end{pmatrix}$, as in SILVEY & TITTERINGTON (1973, Example 4.2, p. 28). Their Example 4.1 (op. cit.) and ATWOOD's (1969, p. 1581) Example 3.1 are of like shape as when $c = \begin{pmatrix}1\\1\end{pmatrix}$, up to a rotation of 45 degrees. ■

Corollary 2 may sometimes be useful in discussing uniqueness of optimal design measures.

Example 3. $\mathfrak{X}=[-1, 0], f(x)=\begin{pmatrix}1\\x\end{pmatrix}, c=\begin{pmatrix}1\\1\end{pmatrix}$, see Fig. 1. Uniqueness of the optimal design $\xi(-1)=1/3$, $\xi(0)=2/3$ for $c'\beta$ may be derived as follows. No singular information matrix is in $\mathfrak{M}(c)$, and non-singularity of M restricts the matrices N of Corollary 2 to have rank 1, by (2.2). The only such N is dd' with $d=\begin{pmatrix}1\\2\end{pmatrix}$. Then $f(x)'\,Nf(x)$ equals 1 if and only if x is -1 or 0, thus determining the support of the optimal ξ. The weights 1/3 and 2/3 are easy to compute. ■

The geometric argument is, of course, also useful when one cannot draw charts as easily as with $k=2$ parameters.

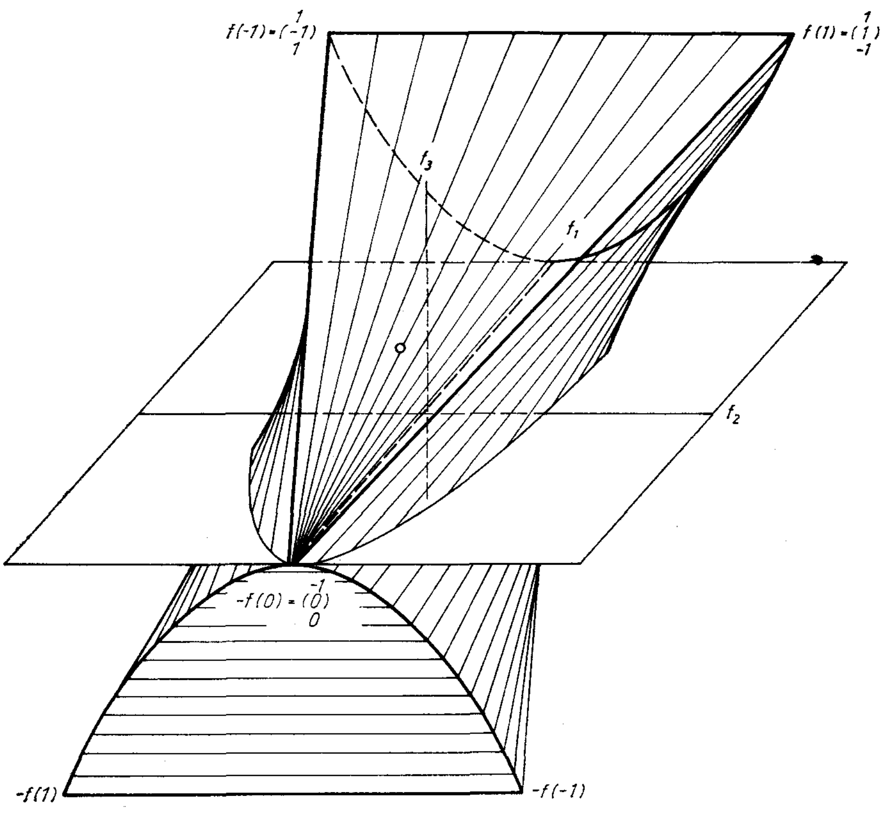

Fig. 2. In Example 4 the regression ball $\mathfrak{R}$ for quadratic regression $f(x)=$ $=(1, x, x^2)'$ on $\mathfrak{X}=[-1, +1]$ is needed. The front arc in the picture is the image $-f(\mathfrak{X})$ since its cross-section with first coordinate fixed is a parabola, the mostly hidden rear arc is $f(\mathfrak{X})$. Once it is seen that the triangle with vertices $-f(0)$, $f(1)$, and $f(-1)$ is a face, its shortest vector $c=(1/5)\,f(1)+(1/5)\,f(-1)-(3/5)\,f(0)$ is a safe candidate for an in-ball radius of $\mathfrak{R}$.

Example 4. $\mathfrak{X}=[-1, +1]$, $f(x)=(1, x, x^2)'$, $c=(-1/5, 0, 2/5)$; see Fig. 2. Since c equals $(1/5)f(1)+(1/5)f(-1)-(3/5)f(0)$, it lies in the face generated by $f(1)$, $f(-1)$, and $-f(0)$, and one has $\varrho(c)=1$. Thus the measure $\xi(0)=3/5$, $\xi(-1)=$ $=\xi(+1)=1/5$ is optimal for $c'\beta$. Note that ξ is also E-optimal for β, as shown by KIEFER (1974, p. 868). ■

It is not true that a design measure which is E-optimal for β is always optimal for some $c'\beta$ where $c\in\mathrm{R}^k$ has Euclidean norm 1. This is shown by the final example which thus disproves Satz 8.3.13 in HUMAK (1977, p. 469).

Example 5. $\mathfrak{X}=[-\pi/2, +\pi/2]$, $f(x)=\begin{pmatrix}\cos x\\ \sin x\end{pmatrix}$. An E-optimal design measure for β is $\xi(+\pi/4)=\xi(-\pi/4)=\frac{1}{2}$, with information matrix $M=\frac{1}{2}I_2$, and optimal value $\lambda_{\max}(M^{-1})=2$. But $r^{-2}=1$, since $\mathfrak{R}$ is the Euclidean unit ball. This discrepancy has nothing to do with multiplicities of eigenvalues, for suppose we change $\mathfrak{X}$ to $\mathfrak{X}(\alpha)=$ $=[-\alpha, +\alpha]$. As α decreases from $\pi/2$ to $\pi/4$ the corresponding quantity $\{r(\alpha)\}^{-2}$ increases continuously from 1 to 2, whereas ξ remains E-optimal for β. ■

Acknowledgement

An initial version of parts of this paper was prepared while the author was a post-doctoral fellow of the Deutsche Forschungsgemeinschaft at Stanford University, and appeared as Technical Report No. 128, Department of Statistics, Stanford University.

References

ATWOOD, C. L. (1969). Optimal and efficient designs of experiments. *Ann. Math. Statist.* **40,** 1570–1602.

BANDEMER, H., u. a. (1977). *Theorie und Anwendung der optimalen Versuchsplanung* I. Akademie-Verlag, Berlin.

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

FEDOROV, V. V. (1972). *Theory of Optimal Experiments.* Academic Press, N. Y.

FEDOROV, V. V. & MALYUTOV, M. B. (1972). Optimal designs in regression problems. *Math. Operationsforsch. Statist.* **3,** 281–308.

HUMAK, K. M. S. (1977). *Statistische Methoden der Modellbildung.* Band I. Akademie-Verlag, Berlin.

KARLIN, S. & STUDDEN, W. J. (1966). Optimal experimental designs. *Ann. Math. Statist.* **37,** 783–815.

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

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2,** 849–879.

KRAFFT, O. (1978). *Lineare statistische Modelle und optimale Versuchspläne.* Vandenhoeck & Ruprecht, Göttingen.

PUKELSHEIM, F. (1980). On linear regression designs which maximize information. *J. Statist. Plann. Inference* (Forthcoming).

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, N. J.

SIBSON, R. (1974). D_A-optimality and duality. pp. 677–692 in vol. II of: GANI, J., SARKADI, K. & VINCZE, I. (eds.). *Progress in Statistics*. European Meeting of Statisticians, Budapest (Hungary) 1972. Colloquia Mathematica Societatis János Bolyai, 9. North Holland Publ. Comp., Amsterdam.

SILVEY, S. D. (1978). Optimal design measures with singular information matrices. *Biometrika* **65**, 553–559.

SILVEY, S. D. & TITTERINGTON, D. M. (1973). A geometric approach to optimal design theory. *Biometrika* **60**, 21–32.

WHITTLE, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc.* B **35**, 123–130.

WYNN, H. P. (1972). Results in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc.* B **34**, 133–147. Discussion of Dr. Wynn's and of Dr. Laycock's papers. Op. cit. 170–186.

Zusammenfassung

Es wird ein kurzer Beweis zur Konstruktion c-optimaler Maße in der Versuchsplanung gegeben, wobei besondere Aufmerksamkeit der diesem Problem innewohnenden Geometrie gewidmet wird. Eine Anzahl von Beispielen und Gegenbeispielen, welche sich auf andere Resultate der Literatur beziehen, wird aufgeführt.

Резюме

Дается короткое доказательство к конструкции c-оптимальных мер в планировании эксперимента. Обращается особое внимание геометрии этой проблемы. Будет приведено число примеров относящихся к результатам литературы.

Received July 1979.