Ann. Inst. Statist. Math.
Vol. 40, No. 4, 821–827 (1988)

OPTIMUM EXPERIMENTAL DESIGN FOR A REGRESSION ON A HYPERCUBE-GENERALIZATION OF HOEL'S RESULT

EWARYST RAFAJŁOWICZ[1] AND WOJCIECH MYSZKA[2]

[1]*Institute of Engineering Cybernetics, Technical University of Wrocław, Wybrzeże St. Wyspiańskiego 27, 50-370 Wrocław, Poland*
[2]*Institute of Material Science and Mechanical Engineering, Technical University of Wrocław, Wybrzeże St. Wyspiańskiego 27, 50-370 Wrocław, Poland*

(Received September 24, 1986; revised September 16, 1987)

Abstract. In the note Hoel's result (1965, *Ann. Math. Statist.*, **36**, 1097–1106) is generalized to a large family of experimental design optimality criterions. Sufficient conditions for optimality criterion are given, which ensure existence of the optimum experimental design measure which is a product of design measures on lower dimensional domains.

Key words and phrases: Experimental design, regression, D-, L-, A-, Q-optimality.

1. Introduction

In 1965 Hoel published the following result (see Hoel (1965*a*)): If the regression function

$$(1.1)\qquad EY(x,y)=\sum_{\alpha=1}^{s}\sum_{\beta=1}^{r}c_{\alpha\beta}g_\alpha(x)h_\beta(y)\,,$$

is estimated from uncorrelated observations in $[0,1]^2$, then there exists a G-optimum experimental design, which is a "product" of G-optimum designs for estimating the regression functions:

$$(1.2)\qquad EG(x)=\sum_{\alpha=1}^{s}a_\alpha g_\alpha(x),\qquad EH(y)=\sum_{\beta=1}^{r}b_\beta h_\beta(y)\,.$$

The definition of G-optimality is recalled in Section 2 for readers' convenience (see also Kiefer (1974), Silvey (1980), Fedorov (1982) and Pazman (1986)). The above result holds if experimental designs are all probability measures on $[0,1]^2$ and $[0,1]$, respectively, and by product of designs we mean product of the corresponding measures.

Applications of this result to analytical construction of G, D-optimum designs are given in Hoel (1965*a*, 1965*b*) and Fedorov (1982). It seems, however, that the advantages of Hoel's decomposition in constructing numerical algorithms have not been fully exploited. As is known (see e.g., Atwood (1978), Wu (1978), Wu and Wynn (1978), Silvey (1980) and Fedorov (1982)), the most difficult and time consuming step in a typical iterative algorithm is that of choosing the next point to be included into a design. Realization of this step, in turn, requires iterative search of extremum of a certain multivariable function. Difficulties in iterative search can be reduced when Hoel's decomposition can be applied to a problem at hand. The above facts motivated us to try to extend Hoel's result to a larger class of optimality criterions, leaving details of computational algorithms outside the scope of this note. In view of recent advances in experiment design theory (Whittle (1973), Kiefer (1974), Ash and Hedayat (1978), Pazman (1980), Silvey (1980) and Fedorov (1982) for example), our task is not too difficult, but the above stated practical motivations seem to justify our algebraic considerations.

We stress that our results are valid for the, so called, asymptotical designs (see Pazman (1986), p. 17) approximating fixed-size designs very well, when the number of observations tends to infinity. This class of experimental designs is commonly used in recent monographs and contributions (Whittle (1973), Kiefer (1974), Ash and Hedayat (1978), Atwood (1978), Wu (1978), Wu and Wynn (1978), Pazman (1980, 1986), Silvey (1980) and Fedorov (1982)).

2. Statement of the problem and the main result

We consider the regression function $EY(x) = a^T \cdot f(x)$ of a special structure, which generalizes (1.1). Namely, the column vector $f(x)$ of continuous and linearly independent functions on a compact set X is of the form:

$$f(x) = g_1(x^{(1)}) \otimes g_2(x^{(2)}) \otimes \cdots \otimes g_r(x^{(r)}) \triangleq \bigotimes_{i=1}^{r} g_i(x^{(i)}) , \tag{2.1}$$

where $x^{(i)}$, $i = 1, 2, \ldots, r$ are subvectors of the vector x. Furthermore, $x^{(i)} \in X_i$, $i = 1, 2, \ldots, r$, where X_i is a compact set and $X = X_1 \times X_2 \times \cdots \times X_r$. In (2.1), $g_i: X_i \to R^{m_i}$ is $m_i \times 1$ vector of continuous and linearly independent functions, while $\otimes$ denotes Kronecker's, or direct, product of matrices (see Marcus and Minc (1964), Lankaster (1969) or Graham (1981) for its definition). The vector of unknown parameters a is an $m \triangleq \prod_{i=1}^{r} m_i$ column vector, which is estimated from uncorrelated observations of $Y(x_1)$, $Y(x_2)$,... with variances σ^2 independent of $x \in X$. Our aim is to choose an experimental design ξ, which is from the class Ξ of all probability measures on X including all discrete measures. This choice is based on the information matrix

$$(2.2) \qquad M(\xi) = \int_X f(x) f^T(x) \xi(dx) ,$$

as well as on an optimality criterion $\Phi(M(\xi); m)$, which is a real valued function of the information matrix M, and its dimension m (see Whittle (1973), Kiefer (1974), Pazman (1980) and Silvey (1980) for discussion). Every design $\xi \in \Xi$, which maximizes $\Phi(M(\xi); m)$, over Ξ is called Φ-optimum. For example, a design $\xi^* \in \Xi$ which maximizes

$$\Phi(M(\xi); m) = -\sigma^2 \max_{x \in X} f^T(x) M^{-1}(\xi) f(x) ,$$

over $\xi \in \Xi$ with $M(\xi)$ nonsingular is called G-optimum experimental design (see Hoel (1965*a*), Kiefer (1974), Ash and Hedayat (1978), Pazman (1980), Silvey (1980) and Fedorov (1982)) for its interpretation and further examples).

Concerning Φ we adopt the following assumptions:

(A.1) Φ is concave and differentiable in M_+, which is the subset of nonsingular matrices in $M \triangleq \{M(\xi): \xi \in \Xi\}$.

Examples of criterions, for which (A.1) holds are given in Kiefer (1974). Let $F(M; m)$ be an $m \times m$ matrix with elements $\partial \Phi(M; m) / \partial m_{ij}$, where m_{ij} are elements of M. Let Ξ_i with elements $\xi_i, \eta_i, \ldots$ be the class of all probability measures on X_i. Note that product measures

$$(2.3) \qquad \xi(dx) = \prod_{i=1}^{r} \xi_i(dx^{(i)}); \qquad \xi_i \in \Xi_i, \quad i = 1, 2, \ldots, r ,$$

form a subset of Ξ, denoted by $\Xi\pi$. For $\xi_i \in \Xi_i$ define $m_i \times m_i$ matrix $M_i(\xi_i) = \int_{X_i} g_i(x^{(i)}) g_i^T(x^{(i)}) \cdot \xi_i(dx^{(i)})$ and let $\hat{\xi}_i \in \Xi_i$ be a measure for which

$$(2.4) \qquad \max_{\xi_i \in \Xi_i} \Phi(M_i(\xi_i); m_i) = \Phi(M_i(\hat{\xi}_i); m_i) .$$

Its existence is guaranteed by compactness of X_i and by continuity of g_i and Φ.

Note that for every $\xi \in \Xi\pi$

$$(2.5) \qquad M(\xi) = \bigotimes_{i=1}^{r} M_i(\xi_i) ,$$

since for Kronecker's product (see Lankaster (1969) and Graham (1981))

$$(A \otimes B) \cdot (C \otimes D) = (A \cdot C) \otimes (B \cdot D)\,, \tag{2.6}$$

where A, B, C and D are matrices of appropriate dimensions. Our aim is to show that Hoel's result extends to criterions with gradient matrices F fulfilling the following conditions

(A.2) If $M = \bigotimes_{i=1}^{r} M_i$, then

$$F(M; m) = \bigotimes_{i=1}^{r} F(M_i, m_i)\,, \tag{2.7}$$

where M, M_i are $m \times m$ and $m_i \times m_i$ matrices, respectively, and $m = \prod_{i=1}^{r} m_i$.

(A.3) If $M \in M_+$ is an $m \times m$ matrix, then $F(M; m)$ is nonnegative definite; $m = 1, 2, \ldots$.

Examples of criterions, for which (A.2), (A.3) hold are given in the next section. We confine our attention to the so called regular optimality criterions (see Pazman (1980) and Silvey (1980)), which assure estimability of all parameters in a.

THEOREM 2.1. *Under* (A.1), (A.2) *and* (A.3) *one can find a Φ-optimum experimental design $\hat{\xi} \in \Xi$, which is of the form*

$$\hat{\xi}(dx) = \prod_{i=1}^{r} \hat{\xi}_i(dx^{(i)})\,, \tag{2.8}$$

where $\hat{\xi}_i$, $i = 1, 2, \ldots, r$ are defined by (2.4).

PROOF. Recall Kiefer (1974) that under (A.1) a design $\tilde{\xi} \in \Xi$ is Φ-optimum iff

$$\max_{x \in X} \Psi(x, \tilde{\xi}; m) = \int_X \Psi(x, \tilde{\xi}; m)\tilde{\xi}(dx)\,, \tag{2.9}$$

where $\Psi(x, \tilde{\xi}; m) \triangleq f^T(x)F(M(\tilde{\xi}); m)f(x)$, $x \in X$. It suffices to show that for (2.8) condition (2.9) holds. From (2.1), (2.5), (2.6), (2.7), (2.8) and (A.2) it follows that

$$\Psi(x, \hat{\xi}; m) = \prod_{i=1}^{r} \Psi_i(x^{(i)}, \hat{\xi}_i; m_i)\,, \tag{2.10}$$

where $\Psi_i(x^{(i)}, \hat{\xi}_i; m_i) \triangleq g_i^T(x^{(i)})F(M_i(\hat{\xi}_i); m_i)g_i(x^{(i)})$. Now, our result follows from (2.10) and (2.4) since applying (2.9) to (2.4) we get

$$\max_{x^{(i)}\in X_i} \Psi_i(x^{(i)}, \hat{\xi}_i; m_i) = \int_{X_i} \Psi_i(x^{(i)}, \hat{\xi}_i; m_i)\hat{\xi}_i(dx^{(i)}) , \tag{2.11}$$

and simultaneously we have $X = X_1 \times X_2 \times \cdots \times X_r$. On the other hand, (A.3) implies that $\Psi_i(x^{(i)}, \hat{\xi}_i; m_i) \geq 0$, $i = 1, 2, \ldots, r$. Collecting these facts we conclude that for $\tilde{\xi} = \hat{\xi}$ both the maximum operation in (2.9) and the multiple integral can be iterated. This finishes the proof.

3. Discussion

Some comments concerning (A.1), (A.2) and (A.3) are in place. One can notice that (A.1) is a standard assumption made in the equivalence theorems in differential form (see Whittle (1973), Kiefer (1974) and Pazman (1980), but notice that we use maximization instead of minimization of Φ). In these papers, one can also find interpretations and formulas for differentiation of criterions presented below in order to indicate that (A.2) and (A.3) hold for important classes of criterions.

(1) From our theory Hoel's result follows, since for D-optimality $\Phi(M; m) = \ln\det M$ for $M \in M_+$ and $F(M; m) = M^{-1}$. Assumption (A.3) clearly holds, while (A.2) follows from the relationships

$$M^{-1} = \left[\bigotimes_{i=1}^{r} M_i \right]^{-1} = \bigotimes_{i=1}^{r} M_i^{-1} , \tag{3.1}$$

provided that M_i, $i = 1, 2, \ldots, r$ are nonsingular. Proof of the fact that for A, B nonsingular $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ can be found in Lankaster (1969) and Graham (1981). Now, Hoel's result follows from the celebrated theorem of Kiefer and Wolfowitz (Kiefer (1974) and Fedorov (1982)) on equivalence of G- and D-optimality criterions for designs from Ξ.

(2) In Fedorov (1982) the class of L-optimality criterions is defined. They are of the form $\Phi(M; m) = -\operatorname{tr}[WM^{-1}]$ for $M \in M_+$, where W is an $m \times m$ nonnegative definite matrix of weights. In this case, (A.3) holds, since $F(M; m) = M^{-1}WM^{-1}$, $M \in M_+$. Suppose that W can be decomposed as follows

$$W = \bigotimes_{i=1}^{r} W_i , \tag{3.2}$$

where W_i, $i = 1, 2, \ldots, r$ are $m_i \times m_i$ nonnegative definite matrices. Then (A.2) holds, since by (2.6) and (3.1) we obtain:

$$M^{-1} W M^{-1} = \bigotimes_{i=1}^{r} M_i^{-1} W_i M_i^{-1} . \tag{3.3}$$

Factorization (3.2) arises in a natural way in the following special cases

(a) A-optimality, where W is the $m \times m$ unit matrix,

(b) Q-optimality, in which $W = \int_X f(x) \cdot f^T(x)\,dx = \prod_{i=1}^{r} \int_{X_i} g_i(x^{(i)}) \cdot g_i^T(x^{(i)})\,dx^{(i)}$.

(c) C-optimality with $C = f(x_0)$, in which $W = f(x_0) \cdot f^T(x_0) = \bigotimes_{i=1}^{r} g_i(x_0^{(i)}) \cdot g_i^T(x_0^{(i)})$, where $x_0 = [x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(r)}]$ is a given point at which $\Phi(M; m) = -f^T(x_0) \cdot M^{-1} \cdot f(x_0)$ is to be maximized.

(3) Consider L_p-norm criterions of the form $\Phi(M; m) = -[(1/m) \cdot \mathrm{tr}\, M^{-p}]^{1/p}$, $M \in M_+$, p being positive integer. As is known, for $p \to 0$ we obtain D-optimality criterion, while for $p = 1$, A-optimality. We are interested mainly in the case $p \to \infty$, which corresponds to E-optimality. For this class of criterions we have $F(M; m) = m^{-1/p} \cdot \mathrm{tr}\, [M^{-p}]^{(1-p)/p} \cdot M^{-(p+1)}$ and (A.3) clearly holds. Also (A.2) is fulfilled, since (2.6) and (3.1) imply $M^{-p} = \bigotimes_{i=1}^{r} M_i^{-p}$ and $\mathrm{tr}\, [M^{-p}] = \prod_{i=1}^{r} \mathrm{tr}\, [M_i^{-p}]$. Summarizing, Hoel's result can be extended to a large class of regular optimality criterions and many examples of particular multivariable designs can be constructed from known one-dimensional examples.

Acknowledgement

The authors would like to express sincere thanks to the anonymous referee for his comments, clarifying the presentation.

REFERENCES

Ash, A. and Hedayat A. (1978). An introduction to design optimality with an overview of the literature, *Comm. Statist. A—Theory Methods*, **7**, 1295–1325.

Atwood, C. L. (1978). Convergent design sequences, for sufficiently regular optimality criteria, *Ann. Statist.*, **4**, 1124–1138.

Fedorov, V. V. (1982). *Theory of Optimal Experiments*, Academic Press, New York.

Graham, A. (1981). *Kronecker Products and Matrix Calculus with Applications*, Horwood, Chichester.

Hoel, P. G. (1965*a*). Minimax designs in two dimensional regression, *Ann. Math. Statist.*, **36**, 1097–1106.

Hoel, P. G. (1965*b*). Optimum designs for polynomial extrapolation, *Ann. Math. Statist.*, **36**, 1483–1493.

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

Lankaster, P. (1969). *Theory of Matrices*, Academic Press, New York.

Marcus, M. and Minc, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, Boston.

Pazman, A. (1980). Some features of the optimal design theory, A Survey, *Math. Operationsforsch. Statist., Ser., Statistics*, **11**, 415–446.

Pazman, A. (1986). *Foundations of Optimum Experimental Designs*, Reidel, Dordrecht.

Silvey, S. D. (1980). *Optimal Design*, Chapman & Hall, London-New York.

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

Wu, C. F. (1978). Some iterative procedures for generating nonsingular optimal designs, *Comm. Statist. A—Theory Methods*, **7**, 1399–1412.

Wu, C. F. and Wynn, H. P. (1978). The convergence of general step-length algorithms for regular optimum design criteria, *Ann. Statist.*, **6**, 1273–1285.