DESIGN OF EXPERIMENTS UNDER CONSTRAINTS

A. Gaivoronski
V. Fedorov

January 1984
WP-84-8

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

This paper was done in collaboration between the System and Decision Sciences Area (SDS) and the Adaptive Resource Policy Project (ARP). It faces the problem of optimal experimental design. This problem arises in adaptive policy making at the stage of estimating a model's parameters. It can be considered as an optimization problem with both objective functions and constraints dependent upon probabilistic measures. Methods for dealing with such problems have recently been developed in SDS. In this paper, these methods are applied to optimal experimental design which allows us to get nontrivial results both in statistics and optimization theory.

Andrzej Wierzbicki
Chairman
System and Decision Sciences Area

CONTENTS

DESIGN OF EXPERIMENTS UNDER CONSTRAINTS

A. Gaivoronski and V. Fedorov

INTRODUCTION

It is a specific feature of applied systems analysis that the organization and implementation of experiments is a very difficult and expensive process. Any change in controllable variables (for instance, in agriculture, health service, economic experiments, etc.) can lead to significant expense or to some kind of loss which cannot be measured in currency units. Therefore, it is necessary to have methods of experimental design which take into account this side of experimental research. These methods were partly developed in the traditional theory of optimal design (see, for example, Fedorov 1972 and Silvey 1980). In the traditional approach it is usually only assumed that controllable variables belong to some given set (so called operability region). In this paper we try to analyze the experimental design problem under more sophisticated constraints.

From the mathematical point of view, we deal with the designing of experiments which are described by a linear regression model:

$$y_i = \vartheta_t^T f(x_i) + \varepsilon_i, \quad i = \overline{1,N} \tag{1}$$

where $f(x)$ is a $(m \times 1)$-vector of a known basic function, x_i describes conditions of the i-th measurement, ϑ is a $(m \times 1)$-vector of unknown parameters, the subscript t stands for the true value of these parameters, i stands for the number of measurements, $y_i \in R^1$ is the result of the i-th measurement, ε_i is the random error with zero mean and the same variance for all measurements which obviously can be chosen equal to 1 by the appropriate scaling, moreover all errors are uncorrelated.

For model (1) it is natural to use the best linear unbiased estimates (see Rao, 1968)

$$\hat{\vartheta} = \overline{M}^{-1} Y \tag{2}$$

where $\overline{M} = \sum_{i=1}^{N} f(x_i) f^T(x_i)$, $Y = \sum_{i=1}^{N} f(x_i) y_i$ and $\overline{M}$ is supposed to be regular. It is well known that the variance matrix (which defines the precision of estimator $\hat{\vartheta}$) equals

$$D(\hat{\vartheta}) = E[(\hat{\vartheta} - \vartheta_t)(\hat{\vartheta} - \vartheta_t)^T] = \overline{M}^{-1} \tag{3}$$

Matrix $\overline{M}$ is called the *information matrix*. It is clear from (2) and (3) that matrix $\overline{M}$ is defined completely by the set $\{x_i\}_1^N$. If in some points x_i there are r_i measurements, then this matrix is defined by the set

$$\xi_N = \{p_i, x_i\}_1^n, \quad \sum_{i=1}^{n} p_i = 1, \quad p_i = r_i / N$$

which is usually called the *design*, and points x_i are its *supporting points*. If one can control or choose the value of x_i, then it is sensible to look for optimal designs.

The design ξ_N^* is optimal if

$$\xi_N^* = Arg \min_{\xi_N} Q[\overline{M}(\xi_N)] \tag{4}$$

where Q is some precision measure; for instance, it can be $|\overline{M}^{-1}|$, $tr\ \overline{M}$ or $tr\ A\overline{M}$ (for details, see Fedorov 1972 and Silvey 1980).

To specify the extremal problem (4) one should describe (or do some suggestion on) the properties of function Q and the admissible set of designs ξ_N. In traditional experimental design theory, this set is defined through constraints on the supporting points: $x \in X \in R^k$, where X is the "operability" region.

The results of this paper are essentially connected with additional constraints. Namely, we suggest that together with the previous constraint, one can deal with the following constraints:

$$\sum_{i=1}^{h} r_i \zeta_\alpha(x_i) \leq c_\alpha N, \quad \alpha=\overline{1,l} \tag{5}$$

In (5), functions $\zeta_\alpha(x)$ describe some losses when a measurement is done at point x, and usually $\zeta_\alpha(x) > 0$.

As in the traditional case, it is convenient to introduce instead of $\overline{M}(\xi_N)$, a normalized information matrix:

$$M(\xi_N) = N^{-1}\overline{M}(\xi_N)$$

and deal with the function

$$\Psi[M(\xi_N)] = Q[N\overline{M}^{-1}(\xi_N)]$$

Using this new notation, the extremal problem (4)-(5) can be presented in the following form:

$$\xi_N^* = Arg \min_{\xi_N} \Psi[M(\xi_N)], \quad \sum_{i=1}^{n} p_i \varphi(x_i) \leq 0, \quad x_i \in X, \quad i=\overline{1,n} \tag{6}$$

where $\varphi(x_i)=[\varphi_1(x_i),\ldots,\varphi_l(x_i)]^T$, $\varphi_\alpha(x_i)=\zeta_\alpha(x_i)-c_\alpha$.

APPROXIMATE OPTIMAL DESIGN

Extremal problem (6) is discrete $(p_i = r_i / N)$ and its solution is quite difficult for any practical situation. But when N is sufficiently large one can hope that a "continuous" design (when p_i is allowed to equal any value between 0 and 1) can be a good approximation of an exact (discrete) design (compare with Fedorov 1972; Silvey 1980). Moreover, it is convenient to describe a design not by the set of "weights" $\{p_i\}_1^n$, but by the arbitrary probabilistic measure $\xi(dx)$ with the supporting set X. Of course, it *can* happen that some optimal design *could* be described by a continuous measure, which is not naturally convenient in practice. But it will be shown later that it is always possible to find a design with the same information matrix, but with a finite number of supporting points.

For continuous design, (6) can be rewritten in the following way:

$$\xi^* = Arg \min_{\xi} \Psi[M(\xi)] \tag{7}$$

$$\Phi(\xi) \le 0 \tag{8}$$

where $\Phi(\xi) = \int_X \varphi(x)\xi(dx)$, and $M(\xi) = \int_X f(x)f^T(x)\xi(dx)$.

In the sequel we shall need fulfillment of the following assumptions:

(a) The set X is compact.

(b) The functions $f(x)$ and $\varphi(x)$ are continuous on X.

(c) $\Psi[M]$ is a convex function.

(d) There exists Q such that $\{\xi: \Psi[M(\xi)] \le Q < \infty, \int_X \varphi(x)\xi(dx) < 0\} = \Xi(Q) \ne \phi$.

(e) For any $\xi \in \Xi(Q)$ and $\bar{\xi} \in \Xi$, where Ξ is the set of designs satisfying to (8),

$$\Psi[(1-\alpha)M(\xi)+\alpha M(\bar{\xi})] = \Psi[M(\xi)]+\alpha\int_X \psi(x,\xi)\bar{\xi}(dx)+\tau(d,\xi,\bar{\xi})$$

where $\tau(d,\xi,\bar{\xi}) = o(\alpha)$.

THEOREM 1. *If conditions (a) and (b) hold, then for any design $\xi \in \Xi$ there can always be found a design $\bar{\xi} \in \Xi$ with the same information matrix $[M(\xi) = M(\bar{\xi})]$, the same value of the cost function $[\int_X \varphi(x)\xi(dx) = \int_X \varphi(x)\bar{\xi}(dx)]$, and containing no more than $\frac{m(m+1)}{2} + l + 1$ supporting points.*

Proof. Since any matrix $M(\xi)$ is symmetric, it is completely described by $m(m+1)/2$ elements. Therefore both $M(\xi)$ and $\Phi(\xi) = \int_X \varphi(x)\xi(dx)$ can be described by a vector of dimension $k = \frac{m(m+1)}{2} + l$. From the definition the set S^* of the corresponding vectors is the convex hull of the set $S = \{q(x), x \in X\} \in R^k$, where $q^T(x) = [f_\alpha(x)f_\beta(x), \alpha \ge \beta, \varphi_\gamma(x)]$, $\alpha,\beta = \overline{1,m}$, $\gamma = \overline{1,l}$. Due to Caratheodory's theorem, any point s^* from S^* can be represented in the form

$$s^* = \sum_{i=1}^{k+1} p_i s_i$$

where $s_i \in S$, $p_i \geq 0$, $\sum_{i=1}^{k+1} p_i = 1$. This fact proves the theorem.

THEOREM 2.

I. If the conditions (a)-(c) hold, then a necessary and sufficient condition for a design ξ^* to be optimal is fulfillment of the inequality

$$\min_{\xi \in \Xi} \int_X \psi(x,\xi^*)\xi(dx) \geq 0 \tag{9}$$

II. The set of optimal designs is convex.

Proof. The inequality (9) follows from assumption (c) and from the fact that a necessary and sufficient condition for M^* to be the solution of the minimization problem $\min\Psi[M]$, where Ψ is a convex function, is the nondecreasing of Ψ along any feasible direction (compare for instance with Whittle 1973 and Fedorov 1981). The convexity of the set of optimal designs is the obvious consequence of the convexity of the function ψ .

Remark. If there are no constraints (8), then

$$\min_{\xi \in \Xi} \int_X \psi(x,\xi^*)\xi(dx) = \min_{x \in X} \psi(x,\xi^*)$$

and Theorem 2 coincides with the well-known "equivalency theorem" from traditional experimental design theory (see, for instance, Fedorov and Malyutov 1972; Kiefer 1974; Whittle 1973).

According to Theorem 2 we should solve problem (9) in order to check particular plans for optimality. This problem is much easier than the initial one because it is linear with respect to ξ. However, it still remains an optimization problem in regard to probabilistic measures and further attempts should be made to reduce it to a more tractable one. This can be done by applying duality results for optimization problems in which the objective function depends on probabilistic meausres (Ermoliev 1970; Ermoliev and Nedeva 1982; Ermoliev, Gaivoronski, and Nedeva 1983).

THEOREM 3.

Suppose that conditions (a)-(c) are held and function is continuous with respect to ξ^*. Then

1. $\min_{\xi \in \Xi} \int \psi(x,\xi^*)\xi(dx) = \max_{u \in U^+} \varphi(u)$ where $U^+ = \{u : u = (U_1, \ldots, U_l),\ u_i \geq 0\};\ \varphi(u) = \min_{x \in X}[\psi(x,\xi^*) + u^T\ \psi(x)]$.

2. *For any $\bar{\xi}$ such that*
$\int \psi(x,\xi^*)\bar{\xi}(dx) = \min_{\xi \in \Xi} \int \psi(x,\xi^*)\xi(dx)$
there exists $\bar{u}$ such that $\varphi(\bar{u}) = \max_{u \in U^+} \varphi(u)$ where $\bar{\xi}$ has a support set belonging to
$X(\bar{u}) = \{x : x \in X,\ \varphi(\bar{u}) = \psi(x,\xi^*) + \bar{u}^T \psi(x)\}$.

3. *Among the solutions of (9) there always exists one with no more than l supporting points.*

This theorem is actually a re-statement of Theorem 1 from a paper by Ermoliev, Gaivoronski, and Nedeva (1983). It reduces problem (9) to a finite-dimensional minimax problem.

Therefore, in Theorem 2 the unequality (9) can be replaced by the following one

$$\max_{u \in U^+} \min_{x \in X} [\psi(x,\xi^*) + u^T \varphi(x)] \geq 0 \tag{10}$$

which is more similar to the "traditional" condition. In the following notation $q(x,u,\xi) = \psi(x,\xi) + u^T\varphi(x)$ will be used.

Let u^* be a solution of (10) and all constraints from (8) are active; i.e.,

$$\int_X \varphi(x)\xi^*(dx) = 0 \tag{11}$$

In the opposite case one can consider (7) which contains fewer, and only active, constraints.

THEOREM 4. *If $\int_{X'} \xi^*(dx) \geq \gamma > 0$, then the function $q(x,u^*,\xi^*)$ achieves zero on the set X'.*

Proof. Let us suggest that at least on some set X':

$$\psi(x,\xi^*) + u^{*T}\varphi(x) \geq \delta > 0 \tag{12}$$

Then, due to (10) and (12):

$$\int_X [\psi(x,\xi^*) + u^{*T}\varphi(x)]\xi^*(dx) \geq \int_{X \setminus X'} [\psi(x,\xi^*) + u^{*T}\varphi(x)]\xi^*(dx) + \int_{X'} \delta\xi^*(dx) \geq \delta\gamma > 0$$

But at the same time

$$\int_X [\psi(x,\xi^*) + u^{*T}\varphi(x)]\xi^*(dx) = \int_X \psi(x,\xi^*)\xi^*(dx) + u^{*T}\int_X \varphi(x)\xi^*(dx) = 0$$

because for any design ξ

$$\int_X \psi(x,\xi)\xi(dx) = 0$$

due to condition (c), and the second summand equals zero due to (11). This contradiction proves the theorem.

Remark. If the design ξ^* contains a finite number of supporting points x_i^*, $i=\overline{1,n}$, then for all of them,

$$q(x_i^*, u^*, \xi^*) = 0$$

Of course Theorems 2 and 3 cannot provide prescriptions for the design's construction in general, but very often they help in the understanding of some essential features of them.

Example 1. Let us consider the design problem for one-dimension polynomial regressions:

$$y_i = \sum_{\alpha=1}^{m} \vartheta_\alpha x^{\alpha-1} + \varepsilon_i \tag{13}$$

with the D-criterion of optimality:

$$\Psi[M] = \ln |M^{-1}| \tag{14}$$

and with the following constraints

$$|x| \leq 1, \quad \int_{-1}^{1} \varphi(x)\xi(dx) \leq 0, \quad \varphi \in R^l .$$

Let us suggest that $\varphi(x)$ are continuous functions on the interval $|x| \leq 1$ and that the system

$$1, x, \ldots, x^{2(m-1)}, \varphi_1(x), \ldots, \varphi_l(x) \tag{15}$$

is a Chebyshev system on the same interval.

It is easy to check that the conditions (a)-(c) are fulfilled and the results of Theorems 2 and 4 take place here. For D-optimal designs one has $\psi(x,\xi) = m - f^T(x) M^{-1} f(x)$ (see, for instance, Fedorov 1972). In our case $f^T(x) = (1, x, \ldots, x^{m-1})$, and therefore

$$q(x,u,\xi) = \psi(x,\xi) + u^T\varphi(x) = m - \sum_{\alpha\beta=1}^{m} M_{\alpha\beta}^{-1}\, x^{\alpha} x^{\beta} + u^T\varphi(x)$$

In other words, the function $q(x,u,\xi)$ is a linear combination of the function (15). It is known that a linear combination with some non-zero coefficients of s functions which is a Chebyshev system can have no more that s roots. Therefore the function $q(x,u,\xi)$ has no more than $(2m+l)$ roots and has no more than $m+\frac{l}{2}$ (if l is even) or $m+\frac{l+1}{2}$ (if l is odd) minima on the interval $|x| \le 1$. But in accordance to Theorem 3, the function $q(x,u^*,\xi^*)$ should approach its low boundary at the supporting points of an optimal design. So *their number cannot exceed* $m+\frac{l}{2}$, *if* l *is even, or* $m+\frac{l+1}{2}$, *if* l *is odd*, which is much less than the upper boundary from Theorem 1.

Example 2. Let us now apply the simplest version of (13), (14) with $m=2$ (simple linear regression), but with the following constraints.

$$x^2 - c \le 0, \qquad |x| \le 1, \qquad c < 1 \tag{16}$$

In this case, system (15) is not a Chebyshev one, and therefore the previous result cannot apply.

According to Theorem 2 and the symmetry of constraints (16), the information matrix $M(\xi)$ for any optimal design should be diagonal. For a diagonal matrix $M(\xi)$ one has:

$$q(x,u,\xi) = 1 - uc + [u - M_{22}^{-1}(\xi)]x^2$$

It is evident that $q(x,u,\xi) \equiv 0$, when $u^* = c^{-1}$ and

$$c = M(\xi^*) = \int_{-1}^{1} x^2 \xi^*(dx) \tag{17}$$

Therefore if a design ξ^* which satisfies (17) can be found, then according to Theorem 2, it will be optimal design. In fact (17) describes a family of distributions with the given second moment and it is not difficult to find some members of it. For instance, the following two designs:

$$\xi_1^* = \begin{Bmatrix} -1, & 0, & +1 \\ c/2, & 1-c, & c/2 \end{Bmatrix}$$

and

$$\xi_2^* = \begin{Bmatrix} -\sqrt{c}, & \sqrt{c} \\ 1/2, & 1/2 \end{Bmatrix}$$

belong to this family. We remember that in the traditional case (only one constraint: $|x| \leq 1$), the optimal design problem has a unique solution:

$$\xi^* = \begin{Bmatrix} -1, & +1 \\ 1/2 & 1/2 \end{Bmatrix}$$

REFERENCES

Ermoliev, Y. 1970. Methods for Stochastic Programming in Randomized Strategies. Vol.1:3-7. Moscow: Kibernetika.

Ermoliev, Y. and C. Nedeva. 1982. Stochastic Optimization Problems with Partially Known Distribution Functions. CP-82-60. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Ermoliev, Y., A. Gaivoronski, and C. Nedeva. 1983. Stochastic Optimazation Problems with Incomplete Information About Distribution Functions. WP-83-13. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Fedorov, V.V. and M.B. Malyutov. 1972. Optimal Designs in Regression Problems. Math. Oper. and Statist. (B) 4:300-324.

Fedorov, V.V. 1972. Theory of Optimal Experiments. New York: Academic Press.

Fedorov, V.V. 1981. Active regression experiments, edited by V.V. Penenko. Mathematical Methods in the Design of Experiments. Pp 19-73. Novosibirsk: Nauka. (In Russian.)

Kiefer, J. 1974. General Equivalence Theory for Optimum Design. Ann. Statist. 2:849-879.

Rao, C.R. 1968. Linear Statistical Inference and Its Applications. New York: John Wiley and Sons, Inc.

Silvey, S.D. 1980. Optimal Design. Chapman and Hall.

Whittle, P. 1973. Some General Points in the Theory of Optimal Experimental Design. JRSS (B) 35:123-130.