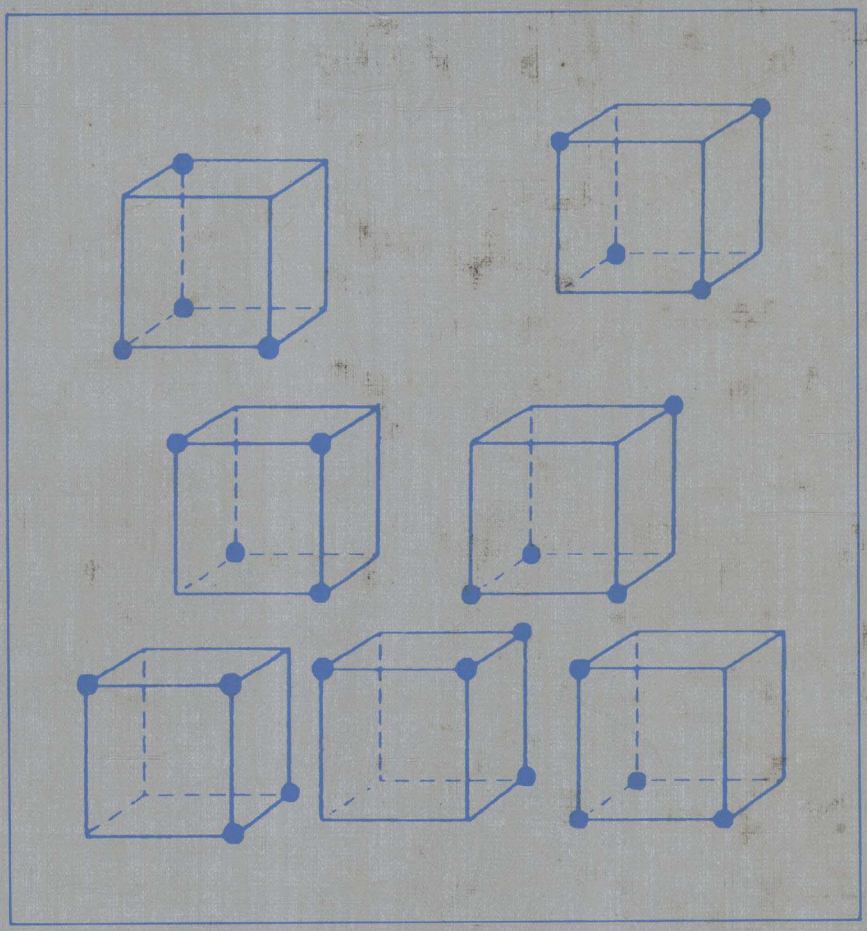
Optimal Design and Analysis of Experiments

Yadolah Dodge
Valeri V. Fedorov
Henry P. Wynn
Editors



North-Holland

OPTIMAL DESIGN AND ANALYSIS OF EXPERIMENTS

edited by

Yadolah DODGE

University of Neuchâtel
Switzerland

Valeri V. FEDOROV

International Institute
for Applied Systems Analysis
Austria

Henry P. WYNN

City University, London
England

1988

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD · TOKYO

ISBN: 0 444 70458 2

Publishers:
ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

Sole distributors for the U.S.A. and Canada:
ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 Vanderbilt Avenue
New York, N.Y. 10017
U.S.A.

LIBRARY OF CONGRESS
Library of Congress Cataloging-in-Publication Data

Optimal design and analysis of experiments / edited by Y. Dodge, V.V. Fedorov, H.P. Wynn.
p. cm.
"Papers presented at the First International Conference-Workshop on Optimal Design and Analysis of Experiments, held in Neuchâtel, Switzerland, from July 25-28, 1988"--Pref.
ISBN 0-444-70458-2
1. Experimental design--Congresses. I. Dodge, Yadolah, 1944- II. Fedorov, V. V. (Valeriĭ Vadimovich) III. Wynn, H. P. IV. International Conference-Workshop on Optimal Design and Analysis of Experiments (1st : 1988 : Neuchâtel, Switzerland)
QA279.O66 1988
519.5--dc19 88-15340
CIP

PRINTED IN THE NETHERLANDS

OPTIMAL DESIGN AND ANALYSIS OF EXPERIMENTS

PREFACE

This book contains invited papers presented at the First International Conference-Workshop on Optimal Design and Analysis of Experiments, held in Neuchâtel, Switzerland, from July 25-28, 1988.

The objective of this book is to present recently developed theories and techniques in optimal design and analysis of experiments along with some related methods such as linear and nonlinear models and quality control.

The papers in this book are grouped in seven sections: (1) Optimal combinatorial designs, (2) Algorithms, (3) Nearest neighbour and cross-over designs, (4) Linear and nonlinear models, (5) Spatial and correlated error models, (6) Quality control, (7) General.

The editors of this volume believe that the volume will be of use to research workers in most branches of applied science. It could also be used as a reference or textbook in universities. The main mathematical prerequisites are matrix algebra, mathematical statistics and some knowledge of statistical inference and optimization theory.

As the editors of this volume, we would like to express our sincere appreciation to the authors whose work and interest made this volume possible. We are very grateful to S. König, N. Rebetez and A.-M. Redard for their careful handling of all correspondence, and to Drs. G. Wanrooy and Ms. Inez van der Heide for handling this volume at Elsevier Science Publishers (North-Holland).

Finally, we wish to express our special thanks to Philip Morris Europe whose contribution made the whole conference-workshop possible, and in particular to Dr. M.C. Bourlas, director of Research and Development, for his invaluable guidance.

University of Neuchâtel
Switzerland
July 1988

The Editors

PROLOGUE

Ho, saki, haste, the beaker bring,
Fill up, and pass it round the ring;
Love seemed at first an easy thing –
But ah! the hard awakening.

Hafiz Shirazi:
Persian Poet (1348-1398)

TABLE OF CONTENTS

Optimal Design and Analysis of Experiments
Y. Dodge, V.V. Fedorov and H.P. Wynn (Editors)

OPTIMAL DESIGN OF EXPERIMENTS : AN OVERVIEW

Y. Dodge
University of Neuchâtel
Switzerland

V.V. Fedorov
International Institute For Applied Systems Analysis
Austria

H.P. Wynn
City University, London
England

A brief introduction to optimal design and analysis of experiments is given. A selected bibliography on the subject is provided.

1. HISTORY

The publication of Jack Kiefer's collected works (Kiefer, 1984) allows us to use the word "classical" for the theory of optimal design up to the beginning of the 1980's. He and co-workers had developed in great detail their early ideas of continuous and exact optimal design theory. Some of the mathematical theory, such as that on weighing designs, was so much in Kiefer's particular style that it will be difficult to surpass. Other material, such as the basic theory and some of the computational ideas, continues to be developed in different directions. He had also begun work on designs for correlated models which is a rather active field at present.

The classical theory of optimal design emerged out of three main historical roots: (1) the combinatorial theory which dates back to R.A. Fisher and F. Yates in the 1920's and 1930's and developed mathematically in many directions under the hands of R.C.Bose and others (2) the use of multi-factor and response surface designs in the 1950's and 1960's by G.E.P.Box, N. Draper and others and (3) the decision theoretic methods initiated by A. Wald. Although the framework occupied by optimal design theory was taken from the first two of these it was the third, the decision theoretic, which generated the main new ideas. For this reason the great burst of activity in mathematical statistics in the 1960's and 1970's included optimal design. During this period the

theory acquired followers in Europe. The United Kingdom contributed results on algorithms simultaneously with Eastern Europe where there was close association with the physical sciences, chemistry and optimization theory. In the F.D.R. and G.D.R. considerable contributions were made to the optimization theory approach also.

Despite the use of computer-based methods by a handful of U.S. and U.K. companies in the 1970's and by research workers in Eastern Europe what had been missing was a serious commitment to industrial collaboration by those on the theoretical side. Even the application of the more industrial response surface methodology had stabalized by around 1980. Experimental design was beginning to take second place to more fashionable areas such as resampling and other new developments in data analysis. Then simultaneously several factors worked to produce the resurgence of interest in the subject which we hope is reflected in the papers in this volume. First, there had been theoretical successes in new areas such as correlated models, second, the computer was beginning to uncover new catalogues of good designs and third, and most important, the industrial importance of the subject has begun to be realized. This latter is due in part to the role of experimental design in the quality improvement revolution which is sweeping the West in a response to its success in Japan. But it is also due to successes in branches of physical science such as the spatial sampling of physical systems and simulation design (computational experiments).

2. OPTIMAL DESIGN, BACKGROUND

The purpose of experimental design is the selection of locations $x_1,\ldots,x_n$ at which to take observations $Y(x_1),\ldots,Y(x_n)$ on some stochastic process $Y(x)$. The usual additive set-up is that

$$\text{observation} = \text{model} + \text{error}$$

which we may write as

$$Y(x) = f(x,\Theta) + \varepsilon(x),$$

where Θ *is a vector of unknown parameters. For a model which is additive in the parameters we write*

$$Y(x) = f(x)'\,\Theta + \varepsilon(x),$$
$$= \sum_{j=1}^{k} f_j(x)\Theta_j + \varepsilon(x).$$

For the vector of observations $Y = (Y(x_1),\ldots,Y(x_n))'$ *we have*

$$Y = X\Theta + \varepsilon,$$

where $\varepsilon = (\varepsilon(x_1),\ldots,\varepsilon(x_n))'$ *and* $X = (f(x_1),\ldots,f(x_n))'$.

If the $\varepsilon(x_i)$ *have zero mean and are uncorrelated with equal variance,* σ^2, *then the usual least squares estimator of the parameter vector* Θ *is*

$$\hat{\Theta} = (X'X)^{-1}X'Y.$$

The information matrix is $X'X$ *and the covariance matrix of* $\hat{\Theta}$ *is*

$$D_n = \sigma^2\,(X'X)^{-1}.$$

Roughly speaking the smaller this matrix is the better is the design: $d_n = \{x_1,\ldots,x_n\}$.

Typically we minimize a particular functional on D_n. *Ignoring* σ^2 *we may define* Φ*-optimality as*

$$\min_{d_n} \text{trace}\,(D_n{}^p)^{1/p} \qquad (0 \le p \ge \infty)$$

which specializes to min det (D_n) *when p=0 (D-optimality), min trace* (D_n) *when p=1 and min maximum eigenvalue of* D_n *when* $p=\infty$.

Another important criterium is based on the fitted response $Y(x) = f(x)'\Theta$ *at a general location* x. *Thus* $\text{var}(Y(x)) = f(x)'D_nf(x)$ *and the criterion is G-optimality :*

$$\min_{d_n} \max_{x} \text{Var}\,(\hat{Y}(x)).$$

The "continuous" theory replaces d_n *by a mesure* ξ *on the* x*-space which can be thought of as mass distribution of observations. In this case* d_n *is replaced by*

$$D(\xi) = M^{-1}(\xi)$$

where

$$M(\xi) = \int f(x)f(x)' \xi (d_x),$$

the moment matrix for the problem. Under suitable conditions on the $f(x)$ and the set of ξ, the set of $M(\xi)$ is convex and the Φ-criteria can, by transformation be made into a convex functions on this "moment space". It is this which allows optimization methods to be used. The celebrated General Equivalence Theorem (GET) of Jack Kiefer and Jacob Wolfowitz shows that D-optimality and G-optimality are equivalent. Many papers cover different special cases where the x-space are cubes, simplices balls and so on and the functions $f_j(x)$ are polynomials, splines, sinusoids and so on.

An excellent review article is Atkinson (1982) and, of course, Kiefer (1984) is compulsory reading. In Russian the handbooks of Ermakov (1983) is very useful and is to be translated into English. Brodski et al (1982) is a useful collection of designs.

3. ALGORITHMS

Computer algorithms for optimal designs take two basic forms. For continuous theory we can improve a measure ξ by augmentation:

$$\xi' = (1-\alpha)\xi_x + \alpha \xi_x$$

where ξ_x puts mass unity at a new support point x and $0 \leq \alpha \leq 1$. By iteration and suitable choice of x and the "step length" α, at each stage, we hope to improve upon $D(\xi)$. Refinements include the use of more than one new x point and deletion of points with low mass.

Algorithms for exact theory start with a design d_n and produce a new design d_n by adding good points and deleting bad points, so-called excursions, while keeping the sample size fixed at the end of every excursion. The theory suffers from the existence of many local optima but in practice convergence is excellent.

The original continuous algorithms are due to Fedorov (see his book, 1972

and monograph in Russian: Fedorov and Uspensky, 1975) and Wynn (1970, 1972). Work published around the same time by Pazman was brought together in his book (Pazman, 1980, English edition Reidel, 1986). There had been previous work in Russian which had synthesized earlier approaches.

The early work on exchange algorithms is due to Mitchell (1974) and his original algorithm is incorporated in some form into the commercially available packages under the name of DETMAX. One of the best versions, with many extensions, is due to Welch (ACED). An excellent comparative account is Cook and Nachtsheim (1980).

Both ACED and recent work of Fedorov contain ideas of constrained D-optimality in which the maximum mass at any point is bounded above. This was presented at a previous workshop (Fedorov, 1987) and resurrects some earlier work of Wynn (1982).

4. OPTIMAL COMBINATORIAL DESIGN

When the independent variables x are qualitative we often use rather different notation. For example with a two-way classification model the observations satisfy

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

where the α_i are row (treatment) parameters and the β_j are column (block) parameters. Interest may be centred on subsets of parameters, or functions there of, such as contrasts $\tau = \Sigma c_i \alpha_j$ where $\Sigma c_i = 0$. With a little effort we can find the special information matrix for a suitably chosen set of τ and apply any of the criteria already mentioned. When we have a special control treatment (say treatment 1) then it is natural to take $\tau_i = \alpha_1 - \alpha_j$ $(j \neq 1)$ as the contrasts of interest. The subject is called design for "treatment with control".

Much beautiful theory lies on the boundary between combinatorial theory and optimality theory which optimal design problems generate. Following the fundamental ideas of Kiefer many names are associated with this work and the results of Cheng (1978 a, b and subsequent papers) led the way.

5. AUTOCORRELATED MODELS

When the process $Y(.)$ is autocorrelated the optimal design problems become more complex and research is proceding in several directions. The models in section 2 can be rewritten as :

$$Y(x) = f(x,\Theta) + \delta(x) + \varepsilon(x),$$

where the additional term $\delta(x)$ is an autocorrelated process in the sense that $cov(\delta(x_1),\delta(x_2)) \neq 0$ (although $\delta(x)$ is independent of $\varepsilon(x)$ which remains as a white noise process). For example $\delta(x)$ may be isotropic so that the above covariance only depends on some distance $|x_1 - x_2|$ between x_1 and x_2. There are two basic motivations (1) to interpolate or extrapolate for $Y(x)$ at unobserved locations x or (2) to estimate the parameter Θ. In weather forecasting, for example, one is surely interested in the location of stations (defined by x-values) to extrapolate the weather elsewhere. Prediction over time means incorporating time into x.

Estimation of the parameters Θ is more usual in the analysis of field experiments in which "uniformity" gives rises to spatial autocorrelation. This has generated a mathematically exciting area of "neighbour designs" that take into account which treatments appear next to which and how often. For example in block design we may count neighbours inside a block if there is autocorrelation in a block.

An important new application is when $\varepsilon(x)$ is missing from the model and $corr(\delta(x_1), \delta(x_2)) = 1$ for x_1 coincident with x_2. The process $\delta(x)$ then serves to represent our ignorance about the unsampled values only, the process being essentially deterministic. This is a good model for simulation design since a given input x should, except for computational error, lead to the same $Y(x)$ on every computer run. Optimal spatial design then provides an alternative to Monte Carlo methods.

The paper by Ylvisaker (1987) includes a review of the whole field of spatial sampling with references to earlier joint work with Sacks and connections with Bayesian experimental design (Chaloner, 1984). Applications to computational experiments are described in Sacks and Schiller (1987). The strong con-

nection with Kriging in earth sciences should be noted.

6. ANALYSIS

There has been steady research into the analysis of experiments which has widened the standard assumptions of section 2.

There is a deeper understanding of the structure of classical analysis of variance models and the relationships between random effect models, randomization and the algebric lattice of models and submodels which may arise in hypothesis testing (Speed and Bailey, 1987, is representative). Some of this work is related to the kind of structure which is available on computer packages such as GLIM, GENSTAT and SAS.

When the structure of the incidence matrix is completely arbitrary for a given design and many observations are missing, the analysis under fixed models for several factors becomes complicated. Such designs can be analysed via the notion of connectedness of the incidence matrix (Dodge 1985). Connected designs with minimum number of observations (minimally connected designs) are of special interest in industrial experiments. Construction, characterization and optimality of such designs are described partially in Dodge and Afsarinejad (1985), Birkes and Dodge (1986) and Dodge (1986).

Non-Gaussian and non-linear models present considerable difficulties to an experimenter. Essentially this is because the value of the criterion depends on the unknown parameter Θ. The solution requires some minimax ideas, a Bayesian approach (a prior distribution on Θ) or a local Taylor series expansion around a preliminary estimator $\hat{\Theta}_o$ (Wu, 1985, Titterington, 1987). An important special case is the linear logistic model which has application to the design of dosage-response experiments.

For non-Gaussian errors different estimators can be used such as L_1 methods, Dodge (1987), and some other robust methods, Hample et al (1986), Huber (1981), Dodge (1984), Rousseeuw (1984).

In terms of the original model we may attempt to model variability essentially by assuming that some of the Θ_j parameters are random and estimating

their variance. Important papers are by Kleffe (1984), and Kleffe and Rao (1987).

7. QUALITY CONTROL

The impact of Japanese methods of off-line quality control derived from the work of Genichi Taguchi has been considerable but is mostly confined to practical considerations. There is a need to extract from the more practical and philosophical ideas criteria which can help to design better experiments. With the notation in section 2 one such criterion is to keep $Y(x)$ on to a target c while minimizing $Var(Y(x))$. Thus if $var(Y(x)) = \sigma^2(x)$ then we have

$$\min_x \sigma^2(x) \quad \text{subject to} \quad E(Y(x)) = c.$$

However, since the model is unknown we must use estimates and obtain the stochastic optimization problem

$$\min \sigma^2(x) \quad \text{subject to} \quad Y(x) = c.$$

But then the solution, say x^*, is itself dependent on the data Y and our design optimality criterion is

$$\min_{d_n} \sigma^2(x^*).$$

The problem is hard and close in spirit to the non-linear model problem and also to quadratic control. The criterion can be extended to other risk functions. In the Taguchi methods these are two addition features (at least). The $\sigma^2(x)$ is artificially generated by varying the observations across noise variables over which the designer has no control, except in an experiment, and there is an attempt to partition x into two sets of variables $x^{(1)}$ and $x^{(2)}$ the first which affects $E(Y(x))$ and the second $\sigma^2(x)$. Two recent papers with discussion are Leon at al (1987) and Box (1988). These new approaches are going to have a large effect on variance estimation raising it to have equal status with the estimation of the mean.