THEORY OF OPTIMAL EXPERIMENTS

V. V. Fedorov

Moscow State University
Interfaculty Laboratory of Statistical Methods
Lenin Hills, Moscow

Translated and Edited by

W. J. Studden and E. M. Klimko

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
LAFAYETTE, INDIANA

ACADEMIC PRESS NEW YORK and LONDON 1972

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DD

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 76-182610

AMS (MOS) 1970 Subject Classification: 62-02

PRINTED IN THE UNITED STATES OF AMERICA

Originally published in Russian under the title
TEORIYA OPTIMAL'NOGO EKSPERIMENTA
by
Izdatel'stvo Moskovskogo Universiteta, 1969.

Contents

Chapter 6—Discriminating Experiments

Chapter 7—Generalized Criteria of Optimality

Introduction

1. At the present level of development of science and technology, many investigations in physics, biology, chemistry, metallurgy, etc., require setting up complicated and expensive experiments. The measurement of any experimental quantity always takes place under the influence of some obstacles which can never be completely eliminated despite the efforts of the researcher to keep them to a minimum. Because of this, the investigator deals not with deterministic, but with random quantities. In some cases the measured quantities are random by their very nature. It is necessary to deal with the measurement of such quantities in quantum mechanics, in biological investigations, in certain problems of chemical kinetics, and in other branches of science.

The necessity of applying the apparatus of mathematical statistics to the reduction of the results of measurements is evident when the random components are commensurate with the results themselves. The corresponding methods of reduction have long been used in experimental practice.

For a long time, the attention of mathematical statistics was focused on the perfection of methods of reduction when the method of conducting the experiment was preestablished. The choice of the

experiment itself, that is, when and where to take measurements, was determined mainly by the intuition of the experimenter. During this time it was necessary to deal with problems which were comparatively simple from the theoretical and experimental viewpoints, and which did not require significant expenditures (financial means, time, limited material resources). Losses related to errors of the intuitive solution for the method of conducting the experiment were not met very often and were not essential.

The development of science and technology led to natural complications in the theoretical interpretation of the results obtained, and in the methods of carrying out necessary experimental investigations. More complicated experimental situations led to sharply increased cost of experimental investigations. As an example, one may cite investigations in the realm of the physics of elementary particles where the necessity of building powerful accelerators makes measurements very expensive. Therefore, the problem of extracting an increased quantity of data from processes under study with finite resources is currently very real. Relying on the intuition of the experimenter for the solution of a given problem becomes less and less hopeful. In connection with this, it is absolutely necessary to give a broad class of methods which would give not only the means of reduction of experimental data, but also would permit the organization of the experiment in an optimal manner.

The mathematical apparatus used in the optimal organization of experiments is based on a composition of mathematical statistics methods and methods of solving extremal problems. Increasingly often, mathematical statistics is necessary for wise construction and elucidation of the basic properties of the criteria of optimality of an experiment. Afterward the problem of optimal organization of an experiment (or more briefly, a design of an experiment) leads to the solution of some extremal problem.

We note that a design is suitable only in those cases when the experimenter clearly sets forth the end-purpose of the investigation being conducted. It may be further added that statistical methods of design are instruments which make attainment of an established goal easier. Moreover, the effectiveness of utilizing any instrument (or apparatus) essentially depends on how well it is used and on the qualifications of the hands using it. In the same way, the effectiveness of applying experimental design methods depends, in the large, on how well they are mastered, and on their appropriate utilization. For example, in

conducting elementary experiments requiring very few resources, it is hardly necessary to apply methods requiring computational resources which could significantly exceed the cost of the experiment itself.

II. Currently, it is possible to divide the mathematical theory of experimental design into two basic areas: design of extremal experiments and design of experiments for an elucidation of the mechanism of a phenomenon. A design of the first type is used in those cases when the experimenter is interested in conditions under which the process being investigated satisfies some criteria of optimality. For example, in the development of a new chemical-technological process, the criteria of optimality consists of requiring maximal output of the products of the reaction. In this case, design consists of finding those values of temperature, pressure of reagents, their percentage of concentration, and so forth, for which the established requirements are satisfied.

Frequently, the experimenter finds it necessary to elucidate the global behavior of an investigated object, or as we shall say in the future, to elucidate the mechanism of a phenomenon. For example in the study of a chemical-technological process it may be necessary to elucidate the dependence of the final products of the reaction on temperature, pressure, reagents, and so forth. In the language of mathematics a similar type of problem is formulated in the following way: It is necessary to find a function which defines the relationship between the end product of the reaction and the quantities introduced at the beginning of the reaction (temperature, percentage of reagent concentration, and so forth). Or in short: Find a mathematical model of the given process. In this way, by an elucidation of the mechanism of a phenomenon, here, in distinction from the usual use of this term, it will be understood to be not a direct investigation of a reaction on the level of elementary particles, molecules and so on, but an investigation of the phenomenological side of an event. In other words, we are indifferent, for example, to the manner in which two molecules react. The important point to us is the dependence of the final product on the percentage of reagent concentration which enters into this reaction and can be directly measured in a given experiment. Having set up an investigation of the mathematical form of the dependence of some quantity on corresponding factors, we will thereby give enough information to the specialist of a branch of science on the basis of which he may call on the necessary theoretical apparatus and be able

to deduce the correct form of the elementary chemical reactions.

The design of an extremal experiment has been adequately investigated. A significant number of articles and surveys have been devoted to an exposition of the corresponding mathematical apparatus. The results of a majority of these are discussed in the book by Nalimov and Chernova [*1*]. The book of Hicks, "Fundamental Concepts in the Design of Experiments" [*2*], should also be pointed out. This book is usually used in a first introduction to the problems under consideration. Along with the broad literature expounding the mathematical aspects of design of extremal experiments there are many works related to the practical application of methods of designing such experiments (cf., for example, [*3*, *4*]).

Essentially little attention in contemporary literature is given to the design of experiments for seeking mathematical models to describe an investigated object. From the Russian literature here we can cite only the monograph of Klepikov and Sokolov [*5*], in which there is a brief section related to the planning of such experiments, and a chapter of Nalimov [*4*]. By comparison with the number of works on extremal design, only a very small collection of works deal with the practical application of methods of experimental design for seeking methamtical models.

This book sets as its aim the presentation, from the viewpoint of practical applications, of the more important and accessible mathematical methods of experimental design, along the line of defining mathematical models which describe an investigated object. For a clear explanation of the possibilities and the region of applicability of each of the methods considered in this volume, theoretical material is accompanied by a significant number of examples.

III. We now consider a more detailed mathematical setup of the problem of experimental design along the line of elucidating the mechanism of a phenomenon. Usually the measured quantity depends on one or several factors, which sometimes we shall call "control variables;" this term attempts to emphasize those values of each of the factors that can be chosen arbitrarily from some given domain. Various quantities can be considered as control variables, depending on the type of experiment. For example: time, scattering angle of particles falling on a target, temperature, tension, output of an experimental apparatus, percentage of reagent concentration in a chemical or biological experiment, and so on.

The column vector

$$x = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{Vmatrix}$$

is set up for each level of these quantities; the coordinates $x_1, x_2, \ldots, x_k$ are equal to the values of the control variables and are enumerated in an order suitable to the experimenter.

The space of dimension k, on which the vector x is defined, is called the factor space or the space of control variables. The collection of points of this space, where measurements are possible (that is, the corresponding values of the control variables $x_1, x_2, \ldots, x_k$ which can be realized by the experimenter) is called the region of possible measurements or the domain of actions, and in this book is indicated by X. Determination of the boundary of the region X plays an important role in the design of optimal experiments. In some cases, these boundaries are defined by the very nature of the control variables. For example, the percentage of reagent concentration cannot be less than 0 or greater than 100%, the dimensions of details under investigation are negative, and so on. In other cases, met significantly more often, the boundary of the domain of action is determined by the characteristics of the apparatus used by the investigator or by the form of the process under study. Indeed, the upper boundary of temperature will be determined either by the power of the heat source or by the thermal insulation properties of the materials. The upper limits for the speed of accelerated elementary particles is determined by the parameters of the accelerator. The reader can extend without difficulty the list of analogous examples, related to real experiments encountered in the branch of science familiar to him.

The problem of setting up an experiment for finding a mathematical model, as already noted, is the search for relations among the measured quantities (sometimes there can be several) and the control variables. Since the results of an observation are random quantities, it makes sense in the majority of cases to talk about the relation of the mean value of the quantity under study to the control variables. In the future we shall assume that this relation can be written by means of some function

$$E(y \mid x) = \eta(x),$$

where $E(y \mid x)$ is the mean value of the quantity y under study for

values of the control variables defined by means of the coordinates of the vextor x. The function $\eta(x)$ depends on unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$, and in the general case, its analytic form can also be unknown.

Beginning the search for a mathematical model [the function $\eta(x)$], the experimenter possesses some prior information. The degree of this information can be characterized at three basic levels.

1. The analytic form of the function $\eta(x) = \eta(x, \theta)$ is known. It is required to determine or estimate the unknown parameters

$$\theta = \begin{Vmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{Vmatrix}.$$

2. It is known that the analytic form of the function is defined by one of the functions

$$\eta(x) = \begin{cases} \eta_1(x, \theta_1), \\ \eta_2(x, \theta_2), \\ \vdots \\ \eta_\nu(x, \theta_\nu). \end{cases}$$

The dimension of the vectors $\theta_1, \theta_2, \ldots, \theta_\nu$ can be different. It is required to determine which of the functions

$$\eta_1(x, \theta_1), \eta_2(x, \theta_2), \ldots, \eta_\nu(x, \theta_\nu)$$

is correct and to find the unknown parameters.

3. The analytic form of the function $\eta(x, \theta)$ is not known. It is known only that the function $\eta(x)$ can be approximated sufficiently well, in the region of interest to the experimenter, by means of a finite series in some system (or systems) of given functions. It is required to find the best description of the function $\eta(x)$.

Although the decomposition 1–3 is sufficiently coarse, it is possible to find examples where the real experimental situation occupies a position between any two of these levels. The levels 1–3 are convenient from the viewpoint of existing methods of designing an experiment, and, at the same time, they describe fairly well the majority of real cases.

The mathematical apparatus of experimental design with prior

knowledge corresponding to level 1 began to develop about in the mid 1950s, and its development has practically been completed. For this case, the development consists of effective methods of statistical and sequential experimental design. By a statistical design of an experiment, here we understand a prior design of an experiment in its entirety. By a sequential design we mean a design of an experiment in stages. That is, one or several measurements is taken; after these measurements are realized a reduction of the obtained data is carried out and planning of a further stage is begun, and so on. For a broad class of functions $\eta(x, \theta)$, a statistically designed experiment consists essentially in making use of prepared tables describing the characteristics of optimal designs.

Methods of designing experiments for seeking the correct model from some given collection of models (level 2) appeared only recently and undoubtedly will still improve both from the conceptual as well as from the computational point of view. In connection with this fact, basic attention is focused here on the more simple and complete methods. For more complicated or less complete methods, from the theoretical point of view, the presentation of the material takes on a more descriptive character. The majority of the methods considered for designing experiments, with prior knowledge corresponding to level 2, are by their very nature sequential. It is to be pointed out that the given methods of planning experiments are more effective when the number of concurring models $\eta_1(x, \theta_1)$, (x, θ_2),..., $\eta_\nu(x, \theta_\nu)$ is small. This is a statement of the fundamental regularity of design in general: The more we know, the more effectively we can plan. Therefore the problem of the experimenter as a specialist in his branch of science, which gives rise to the necessity of conducting a given investigation, is to seek the smallest possible collection of models based on a careful analysis of available theoretical and experimental data.

A more difficult, less worked out, and very often met problem is the design of experiments when the analytical function $\eta(x, \theta)$ is completely unknown (level 3). Generally, it is hardly possible to design an experiment which would permit the formulation of the problem in its entirety. Nevertheless its solution can be reduced to some sequential procedure, which contains alternating experiments (design and practical realization) of the following form:

1. The functional form $\eta = \eta(x, \theta)$ is known. It is required to determine or estimate the parameter θ.

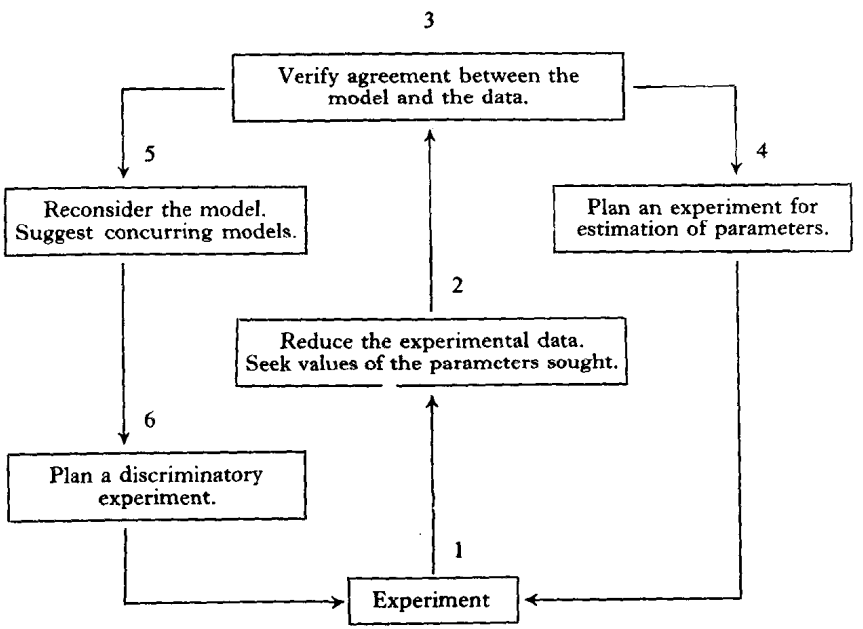
2. On the basis of a theoretical analysis of the occurring hypotheses, or from the results of previous experiments, two or several hypotheses are suggested about the form $\eta(x)$:

$$\eta(x) = \begin{cases} \eta_1(x, \theta_1), \\ \eta_2(x, \theta_2). \end{cases}$$

It is required to find the dependence $\eta_j(x, \theta_j)$; this means the best form describing the object under study.

A more detailed sequential process for seeking a mathematical model is presented in Diagram 1.

Diagram 1



Block 1 corresponds to the experimental stage of the work, i.e., the technical carrying out of previously designed experiments. Usually the carrying out of designed experiments is preceded by conducting some "preliminary" experiment which offers the experimenter rough information about the process under investigation, since in the complete absence of prior information, design is impossible.

The next stage of the work (Block 2) is the computation of estimates

of the parameters θ under the assumption that the functional form of $\eta(x, \theta)$ is known. Sometimes the computation of the estimates of the parameters is preceded by an analysis of experimental data from the point of view of discrimination of concurring models.

After estimates of the parameters are found, it is necessary to verify whether or not the behavior of the function $\bar{\eta}(x) = \eta(x, \bar{\theta})$, where $\bar{\theta}$ is the value of the estimate, agrees with the experimental data. (Block 3).

If $\bar{\eta}(x)$ satisfies the experimental data sufficiently well, then, depending on the circumstances of the experiment, the experimenter must either stop or design a supplementary experiment for estimating the entire collection of parameters or some group of them which are of more interest to the experimenter (Block 4).

If $\bar{\eta}(x)$ does not satisfy the experimental data, then the necessity of a more careful analysis of the occurring phenomena arises. In the absence of any reduction in the number of suggested models we must turn to the design of a more precise experiment. If we have facts which point out the possibility of describing the phenomena under study by some other model in comparison with the original model, then it is necessary to begin to design an experiment which would permit us to clarify which of these models best describes the objects under study (Block 6). In this way, the strategy for conducting an experiment to elucidate the mathematical model, with prior knowledge corresponding to level 3, can be represented in the form of a sequence of cycles 4–1–2–3– and 5–6–1–2–3 (cf. Diagram 1). The order of alteration of these cycles is determined by the results of verifying agreement between the model and the data (Block 3). In many cases, an analogous strategy can be applied to experimental design corresponding to the second level of prior knowledge.

IV. We now go to a brief description of the contents of the following chapters. In Chapter 1, a survey of regression analysis is presented at a level necessary for the construction of the mathematical apparatus of experimental design. The concept of best linear estimator (or best quasi-linear estimator, for the case of functions $\eta(x, \theta)$ depending nonlinearly on the parameters θ) is presented as fundamental in constructing the scheme of regression analysis.

Such an approach permits us to develop the theory of optimal experiments for determining the estimator of an unknown parameter not depending on a concrete form of the distribution function of

the observations. The first section of this chapter has the character of a review, and contains a survey of matrix algebra. Its aim is to gather in one place the basic formulas of matrix algebra, necessary for regression analysis and design of experiments, which are scattered throughout various sources, and at the same time, make it easy for the reader to become acquainted with the basic material. The concluding section of the first chapter is devoted to the formulation of basic optimality criteria of experiments for determining and estimating unknown parameters. The formal definition of an experimental design is introduced there.

Chapter 2 is devoted to an exposition of the basic properties of continuous statistically optimal designs for the various criteria of optimal design and for various criteria of optimality. The analytical and computational methods of constructing such designs is investigated here. A table of characteristics of optimal designs is also presented.

The properties and methods of construction of optimal designs, taking into account the discrete character of the expenditures for carrying out real experiments, is studied in Chapter 3.

In Chapter 4, results are presented for sequential methods of designing experiments in the determination and estimation of unknown parameters. These methods, above all, give perspective to the nonlinear parametrization functions $\eta(x, \theta)$ and in the highest degree satisfy the spirit of Diagram 1.

Chapter 5 generalizes the results obtained in the previous chapters to the case where it is possible to have simultaneous measurements, some of which, generally speaking, are correlated.

Chapter 6 contains information on the discrimination of statistical hypotheses, as they apply to problems of seeking the correct mathematical model, and on various methods of designing discriminating experiments. In contrast to the results of the preceding five chapters, Chapter 6 depends essentially on the analytic form of the distribution function of the observations.

In the concluding chapter, methods of experimental design depending on generalized criteria of optimality are considered which succeed in simultaneously solving the problems of deciding on the correct mathematical model and estimating unknown parameters. Basic attention is focused on criteria using the entropy measure of information.

V. The material is presented in such a way that the basic theoretical results are formulated in the form of theorems. Besides making it possible to concentrate on the most important results, theorems permit the reader to become acquainted with the mathematical methods of designing optimal experiments. He may confine himself, at a first reading, only to the analysis contained in the theorems and explanations, skipping the more complicated and cumbersome proofs. Theorems and lemmas are enumerated in the following form: The first digit is the chapter number, the second digit is the section number, and the last digit is the ordering number in the given section; formulas are handled by means of analogous numbering.

1

Regression Analysis and Optimality Criteria for Regression Experiments

1.1. Basic Elements of Matrix Algebra

This section is of a review character. Detailed illustrations, definitions, and proofs of the basic results presented in what follows can be found in [6–8].

I. Basic properties of matrices.

DEFINITION 1. The rectangular array of numbers

$$\begin{Vmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{Vmatrix} \tag{1.1.1}$$

is called a matrix.

If $m = n$, then the matrix is called square, and the number m, equal to n, is called its order. If $m = 1$ or $n = 1$, then the matrix is respectively a row vector or a column vector. In general, matrices are called rectangular (with dimension $m \times n$) or $m \times n$ matrices. The numbers comprising the matrix are called its elements.