Mathematics Monograph Series 37

Theory and Application of Uniform Experimental Designs

Kai-Tai Fang（方开泰） Min-Qian Liu（刘民千）
Hong Qin（覃 红） Yong-Dao Zhou（周永道）

（均匀试验设计的理论和应用）

SCIENCE PRESS
Beijing

Springer

Mathematics Monograph Series 37

Theory and Application of Uniform Experimental Designs

(均匀试验设计的理论和应用)

Kai-Tai Fang (方开泰) · Min-Qian Liu (刘民千)
Hong Qin(覃红) · Yong-Dao Zhou(周永道)

Science Press
Beijing

Springer

Kai-Tai Fang
Beijing Normal University-Hong Kong Baptist University United International College
Zhuhai, Guangdong, China

and

Institute of Applied Mathematics
Chinese Academy of Sciences
Beijing, China

Min-Qian Liu
School of Statistics and Data Science
Nankai University
Tianjin, China

Hong Qin
Faculty of Mathematics and Statistics
Central China Normal University
Wuhan, Hubei, China

Yong-Dao Zhou
School of Statistics and Data Science
Nankai University
Tianjin, China

ISBN 978-7-03-059110-4
Science Press, Beijing, China

Jointly published with Springer Nature Singapore Pte Ltd.

The print edition is only for sale in Mainland China. Customers from Mainland China please order the print book from: Science Press.

Published by Science Press
16 Donghuangchenggen North Street
Beijing 100717,P.R.China
Printed in Beijing

Foreword

Experiment is essential to scientific and industrial areas. How do we conduct experiments so as to lessen the number of trials while still achieving effective results? In order to solve this frequently encountered problem, there exists a special technique called experimental design. The better the design, the more effective the results.

In the 1960s, Prof. Loo-Keng Hua introduced J. Kiefer's method, the "golden ratio optimization method," in China, also known as the Fibonacci method. This method and orthogonal design which were popularly used in industry promoted by Chinese mathematical statisticians are the two types of experimental designs. After these methods became popular, many technicians and scientists used them and made a series of achievements, resulting in huge social and economic benefits. With the development of science and technology, these two methods were not enough. The golden ratio optimization method is the best method to deal with a single variable, assuming the real problem has only one interesting factor. However, this situation is almost impossible. This is why we only consider one most important factor and fix the others. Therefore, the golden ratio optimization method is not a very accurate approximation method. Orthogonal design is based on Latin square theory and group theory and can be used to do multifactor experiments. Consequently, the number of trials is greatly reduced for all combinations of different levels of factors. However, for some industrial or expensive scientific experiments, the number of trials is still too high and cannot be facilitated.

In 1978, due to the need for missile designs, a military unit proposed a five-factor experiment, where the level of every factor should be higher than 18 and the total number of trials should be not larger than 50. Neither the golden ratio optimization method nor orthogonal design could be applied. Several years before 1978, Prof. Kai-Tai Fang asked me about an approximate calculation of a multiple integration problem. I introduced him to use the number-theoretical methods for solving that problem, which inspired him to think of using number-theoretical methods in the design of the problem. After a few months of research, we put forward a new type of experimental designs that is known as uniform design. This method had been successfully applied to the design of missiles. After our article was published in the early 1980s, uniform design has been widely applied in China and has resulted in a series of gratifying achievements.

Uniform design belongs to the quasi-Monte Carlo methods or number-theoretical methods, developed for over 60 years. When the calculation of a single variable problem (the original problem) is generalized to a multivariable problem, the calculation complexity is often related to the number of variables. Even if computational technology advances greatly, this method is still impossible in application. Ulam and Von Neumann proposed the Monte Carlo method (i.e., statistical stimulation) in the 1950s. The general idea of this method is to put an analysis problem into a probability problem with the same solution and then use a statistical simulation to deal with the latter. This solves some difficult analysis, including the approximate calculation of multiple definite integrals. The key to the Monte Carlo method is to find a set of random numbers to serve as a statistical simulation sample. Thus, the accuracy of this method lies in the uniformity and independence of random numbers.

In the late 1950s, some mathematicians tried to use deterministic methods to find evenly distributed points in space in order to replace the random numbers used in the Monte Carlo method. The set of points that had been found was by using number theory. According to the measure defined by Weyl, the uniformity (of a uniform design) is good, but the independence is relatively poor. By using these points to replace the random numbers used in the Monte Carlo method, we usually get more precise results. This kind of method is called a quasi-Monte Carlo method, or the number-theoretical method. Mathematicians successfully applied this method into approximate numerical calculations for multiple integrals.

In statistics, pseudo-random numbers can be regarded as representative points of the uniform distribution (in cubed units). Numerical integration requires a large sample, but uniform design just uses small samples. Since the sample is more uniform than orthogonal designs, it is preferred for settling the experiment. Of course, when seeking a small sample, the method of seeking a large sample can be used as a reference.

Uniform design is only one of the applications of the number-theoretical method, which is also widely used in other areas, such as the establishment of multiple interpolation formulas, the approximate solutions of systems of some integrals or differential equations, the global extremes of the functions, the approximate representation points for some multivariate distributions, and some problems for statistical inference, such as multivariate normality test and the sphericity test.

When the Monte Carlo method was first discovered in the late 1950s, Prof. Loo-Keng Hua initiated and led a study of this method in China. Loo-Keng Hua and his pioneering results were summarized in our monograph titled "Applications of Number Theory to Numerical Analysis" published in Springer-Verlag Science Press in 1981. These results are one of the important backgrounds and reference materials for my work with Prof. Kai-Tai Fang.

I have worked with Prof. Kai-Tai Fang for nearly 40 years. As a mathematician and a statistician with long-term valuable experience in popularizing mathematical statistics in Chinese industrial sector, he has excellent insight and experience in applied mathematics. He always provided valuable research questions and possible ways to solve the problem in a timely manner. Our cooperation has been pleasant and fruitful, and the results were summarized in our monograph "Number-Theoretic Methods in Statistics" published by Chapman and Hall in 1994.

This book focuses on the theory and application of uniform designs, but also includes many latest results in the past 20 years. I strongly believe that this book will be important for further development and application of uniform designs. I would like to take this opportunity to wish the book success.

Beijing, China

Yuan Wang
Academician of Chinese Academy
of Sciences

Preface

The purpose of this book is to introduce theory, methodology, and applications of the *Uniform experimental design*. The uniform experimental design can be regarded as a fractional factorial design with model uncertainty, a space-filling design for computer experiments, a robust design against the model specification, a supersaturated design and can be applied to experiments with mixtures. The book provides necessary knowledge for the reader who is interested in developing theory of the uniform experimental design.

The *experimental design* is extremely useful in multifactor experiments and has played an important role in industry, high tech, sciences and various fields. Experimental design is a branch of statistics with a long history. It involves rich methodologies and various designs. Comprehensive reviews for various kinds of designs can be found in *Handbook of Statistics, Vol. 13*, edited by S. Ghosh and C. R. Rao.

Most of the traditional experimental designs, like fractional factorial designs and optimum designs, have their own statistical models. The model for a factorial plan wants to estimate the main effects of the factors and some interactions among the factors. The optimum design considers a regression model with some unknown parameters to be estimated. However, the experimenter may not know the underlying model in many case studies. How to choose experimental points on the domain when the underlying model is unknown is a challenging problem. The natural idea is to spread experimental points uniformly distributed on the domain. A design that chooses experimental points uniformly scattered on the domain is called *uniform experimental design* or *uniform design* for simplicity. The uniform design was proposed in 1980 by Fang and Wang (Fang 1980; Wang and Fang 1981) and has been widely used for thousands of industrial experiments with model unknown.

Computer experiments are for simulations of physical phenomena which are governed by a set of equations including linear, nonlinear, ordinary, and partial differential equations or by several softwares. There is no analytic formula to describe the phenomena. The so-called space-filling design becomes a key part of computer simulation. In fact, the uniform design is one of the space-filling designs.

Computer experiment is a hot topic in the past decades. It involves two parts: design and modeling. The book focuses on the theory of construction of uniform designs and connections among the uniform design, orthogonal array, combinatorial design, supersaturated design, and experiments with mixtures. There are many useful techniques in the literature, such as polynomial regression models, Kriging models, wavelets, Bayesian approaches, neural networks as well as various methods for variable selection. This book gives a brief introduction to some of these methods; the reader can refer to Fang et al. (2006) for details of these methods.

There are many other space-filling designs among which the Latin hypercube sampling has been widely used. Santner et al. (2003) and Fang et al. (2006) give the details of the Latin hypercube sampling.

The book involves eight chapters. Chapter 1 gives an introduction to various experiments and their models. The reader can easily understand the key idea and method of the uniform experimental design from a demo experiment. Many basic concepts are also reviewed. Chapter 2 concerns with various measures of uniformity and introduces their definitions, computational formula, and properties. Many useful lower bounds are derived. There are two chapters for the construction of uniform designs. Chapter 3 focuses on the deterministic approach while Chap. 4 on numerical optimization approach. Various useful modeling techniques are briefly recommended in Chap. 5. The uniformity has played an important role not only for construction of uniform designs, but also for many other designs such as factorial plans, block designs, and supersaturated designs. Chapters 6 and 7 present a detailed description on the usefulness of the uniformity. Chapter 8 introduces design and modeling for experiments with mixtures.

The book can be used as a textbook for postgraduate level and as a reference book for scientists and engineers who have been implementing experiments often. We have taught partial contents of the book for our undergraduate students and our postgraduate students.

We sincerely thank our coauthors for their significant contribution to the development of the uniform design, who are Profs. Yuan Wang in the Chinese Academy of Science, Fred Hickernell in the Illinois Institute of Technology, Dennis K. J. Lin in the Pennsylvania State University, R. Mukerjee in Indian Institute of Management Calcatta, P. Winker in Justus-Liebig-Universität Giessen, C. X. Ma in the State University of New York at Buffalo, H. Xu in University of California, Los Angeles, and K. Chatterjee in Visva-Bharati University. Many thanks to Profs. Z. H. Yang, R. C. Zhang, J. X. Yin, R. Z. Li, L. Y. Chan, J. X. Pan, R. X. Yue, M. Y. Xie, Y. Tang, G. N. Ge, Y. Z. Liang, E. Liski, G. L. Tian, J. H. Ning, J. F. Yang, F. S. Sun, A. J. Zhang, Z. J. Ou, and A. M. Elsawah for successful collaboration and their encouragement. We particularly thank Prof. K. Chatterjee who spent so much time to read our manuscript and to give valuable useful comments.

The first author would thank several Hong Kong UGC research grants, BNU-HKBU UIC grant R201409, and the Zhuhai Premier Discipline Grant for partial support. The second author would thank the National Natural Science Foundation of China (Grant Nos. 11431006 and 11771220), National Ten

Thousand Talents Program, Tianjin Development Program for Innovation and Entrepreneurship, and Tianjin "131" Talents Program. The third author would thank the National Natural Science Foundation of China (Grant Nos. 11271147 and 11471136) and the self-determined research funds of CCNU from the college's basic research and operation of MOE (CCNU16A02012 and CCNU16JYKX013). The last author would thank the National Natural Science Foundation of China (Grant Nos. 11471229 and 11871288) and Fundamental Research Funds for the Central Universities (2013SCU04A43). The authorship is listed in alphabetic order.

Zhuhai/Beijing, China	Kai-Tai Fang
Tianjin, China	Min-Qian Liu
Wuhan, China	Hong Qin
Tianjin, China	Yong-Dao Zhou

References

Fang, K.T.: The uniform design: application of number-theoretic methods in experimental design. Acta Math. Appl. Sin. **3**, 363–372 (1980)

Fang, K.T., Li, R., Sudjianto, A.: Design and Modeling for Computer Experiments. Chapman and Hall/CRC, New York (2006)

Santner, T.J., Williams, B.J., Notz, W.I.: The Design and Analysis of Computer Experiments. Springer, New York (2003)

Wang, Y., Fang, K.T.: A note on uniform distribution and experimental design. Chin. Sci. Bull. **26**, 485–489 (1981)

Contents

Chapter 1
Introduction

Experimental design is an important branch of statistics. This chapter concerns with experiments in various fields and indicates their importance, purpose, type of experiments, statistical models, and related designs. Section 1.1 demonstrates several experiments for different purposes and characteristics. This section also presents discussion on two popular types of experiments: (1) physical experiments and (2) computer experiments. Basic terminologies used in experimental design are introduced in Sect. 1.2. Various kinds of experimental designs based on different kinds of statistical models are introduced in Sect. 1.3. They involve the factorial design under ANOVA model, the optimum design under linear regression model, and the uniform design under model uncertainty (or nonparametric regression model). There are many criteria for assessing fractional factorial designs, among which the minimum aberration criterion has been widely used. Section 1.4 gives a brief introduction to this concept and its extensions. Section 1.5 shows the implementation of the uniform design for a multifactor experiment. Readers are recommended to read this chapter carefully so that they can understand the methodology of uniform design and will easily follow the remaining contents of the book.

1.1 Experiments

Scientific experiments are of essential importance for exploring nature. Experiments are performed almost everywhere, usually the purpose of discovering something about a particular process/system. Experiments are often implemented in agriculture, industry, natural sciences, and high-tech. The purpose of an experiment in industrial and chemical engineering is

- To increase process yields;
- To improve the quality of the products, such as to reduce variability and increase reliability;
- To reduce the development time; or/and
- To reduce the overall costs.

In natural sciences and high-tech, the purpose of an experiment would be accommodated in different tasks:

- To evaluate the material alternatives;
- To screen and select the design parameters;
- To determine the values of the key product parameters which have an impact on product performance;
- To evaluate the effects and the interactions of the factors; or/and
- To explore the relationships between factors and responses.

How to find a good design for a specific experiment is an important research area in statistics as most experiments involve random errors. Design and modeling for experiments are a branch of statistics and have been playing an important role in the development of sciences and new techniques.

1.1.1 Examples

Let us present some motivating examples of experiments. We omit the details in the following experiments so that readers may concentrate on the problems and related methodology we are going to introduce.

Example 1.1.1 In a chemical experiment, the experimenter wishes to explore the relationship between the composition of a chemical material (x) and its strength (y) by an experiment. Suppose that the underlying relationship shown in Fig. 1.1 is unknown. How does one design an experiment to find an approximate model (or metamodel) to describe the desired relationship. In this experiment, the chemical composition is called a *factor* and the strength is called a *response*. A natural idea for this experiment is to choose several values of the composition, $x_1, \ldots, x_n$ say, to conduct experiments at these values, and measure the corresponding strengths, denoted by $y_1, \ldots, y_n$. Various modeling techniques applied to the data $(x_1, y_1), \ldots, (x_n, y_n)$ can result in different metamodels, among which the experimenter can choose a suitable one. The choice of experimental points, $x_1, \ldots, x_n$, and modeling techniques are important issues.

Example 1.1.2 This experiment is a typical situation in chemical engineering encountered by the first author in Nanjing in 1972. For increasing the yield (y), three controllable variables varied for study. They are

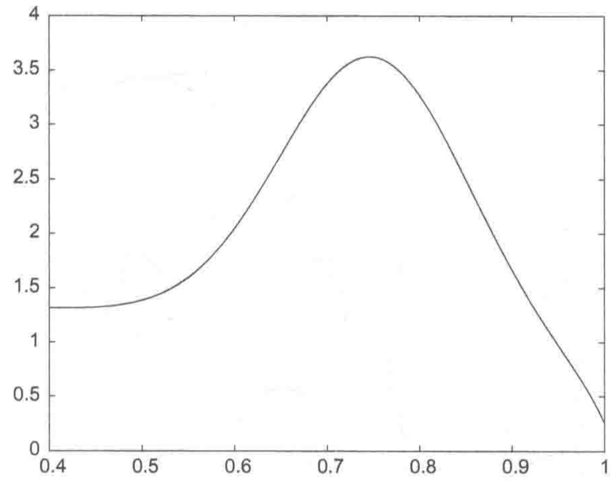

Fig. 1.1 Underlying nonlinear model

A: The temperature of the reaction;
B: The time allowed for the reaction; and
C: The alkali percentage.

The varied intervals of these variables are chosen to be

$$A : 80\,^{\circ}\mathrm{C}\text{–}90\,^{\circ}\mathrm{C};\ B : 90\,\mathrm{min}\text{–}150\,\mathrm{min};\ C\text{: } 5\%\text{–}7\%,$$

respectively. From experience, it is known that there are no interactions among the three factors. There may have a nonlinear relationship between the yield y and the factors. How do we find a design for this experiment? In the literature, a factorial experiment for multiple factors is recommended. Sections 1.3.1 and 1.3.2 will introduce this kind of designs.

Example 1.1.3 This is a real case study introduced by Fang and Chan (2006). Accelerated stress testing is an important method in studying the lifetime of systems. As a result of advancement in technology, the lifetimes of products are increasing, and as new products emerge quickly, their life cycles are decreasing. Manufacturers need to quickly determine the lifetimes of new products and launch them into the market before another new generation of products emerges. In many cases, it is not viable to determine the lifetimes of products by testing them under normal operating conditions. Instead, accelerated stress testing is commonly used, in which products are tested under high-stress physical conditions. The median times to failure of the products are extrapolated from the data obtained using lifetime models. Many different models, such as the Arrhenius model, inverse power rule model, the proportional hazards model, have been proposed based on physical or statistical considerations. Readers may refer to Elsayed (1996) for an introduction of accelerated stress testing. An experimental design is needed to choose the environmental parameters of the accelerated stress test. Three factors are considered as voltage V (Volts), temperature T (Kelvin), and relative humidity H (%). The response is its median time to failure t that is given by

$$t = aV^{-b}e^{c/T}e^{-dH},$$

where a, b, c, d are known constants to be determined. The median time to failure t_0 of an electronic device under the normal operating condition has to be determined under accelerated stress testing. The experimenter wants to explore the model via an experiment. The uniform design was used for this case study.

Example 1.1.4 In an environmental study, an experimenter wishes to conduct a quantitative risk assessment of toxic chemicals present and their interactions. Six chemicals are considered: cadmium (Cd), copper (Cu), zinc (Zn), nickel (Ni), chromium (Cr), and lead (Pb). The experimenter varies the concentration of each chemical in the experiment in order to determine how the concentration affects toxicity. Unfortunately, the underlying model between the response and the six chemical concentrations is unknown. One wants to find a metamodel to the true one by an experiment. Clearly, the range for each chemical concentration should be substantial. The experimenter might choose the following concentrations for each chemical

0.01, 0.05, 0.1, 0.2, 0.4, 0.8, 1, 2, 4, 5, 8, 10, 12, 14, 16, 18, and 20.

Given these 17 levels, there are $17^6 = 24, 137, 569$, almost 24 million, concentration-combinations! It is impossible to conduct an experiment for each concentration-combination. A good experimental design can choose a small number of representative concentration-combinations that still yield a reliable result. Fang and Wang (1994) and Fang et al. (2006) discussed the issues of design and modeling for this experiment in details.

Example 1.1.5 The reversible chemical reaction is a class of important basic reactions in chemistry and chemical engineering. Traditionally, chemists used the deterministic methods to obtain the kinetic rate constants, according to the characteristic of a chemical reaction. Chemists often used the techniques that make a large excess of some reactants involved in reaction, so that their concentration changes can be negligible. Thus, simple relation among the reactants and products is obtained. The chemical kinetics is modeled by a linear system of 11 differential equations:

$$h_j(x, t) = g_j(\eta, x, t), \; j = 1, \ldots, 11, \tag{1.1.1}$$

where x is a set of rate constants, the inputs to the system. A solution to (1.1.1) can be obtained numerically for any input x by the use of 11 differential equations solver, yielding concentrations of five chemical species at a reaction time of 7×10^{-4} seconds. One might be interested in finding a closed-form approximate model that is much simpler than the original one. Atkinson et al. (1998) discussed the possibility of applying D-optimal designs to the kinetics of reversible chemical reaction.

Example 1.1.6 Many products are formed by mixing two or more ingredients together. For making a coffee cake, the ingredients are: X_1 (flour), X_2 (water), X_3 (sugar), X_4 (vegetable shortening), X_5 (flaked coconut), X_6 (salt), X_7 (yeast), X_8 (emulsifiers), X_9 (calcium propionate), X_{10} (coffee powder), and X_{11} (liquid