Journal of the Korean Statistical Society 45 (2016) 610–622

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

An effective approach for the optimum addition of runs to three-level uniform designs

A.M. Elsawah [a,b,c,*], Hong Qin [b]

[a] Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt
[b] Faculty of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China
[c] Division of Science and Technology, BNU-HKBU United International College, Zhuhai 519085, China

ARTICLE INFO

Article history:
Received 15 January 2016
Accepted 7 May 2016
Available online 28 May 2016

AMS 2000 subject classifications:
62K05
62K15

Keywords:
Balanced design
Extended design
Nearly balanced design
Optimal extended design
Uniform design

ABSTRACT

Suppose that an experimenter begins the experimentation using three-level uniform designs. After the experiment is over or during the experimentation, some additional resources become available and the experimenter can afford more runs to the design. The full design obtained by augmenting the additional runs to those of the original design is called as extended design. The extended designs have been applied into computer experiments, microarray experiments and numerical integration. How does the experimenter pick the additional runs and augment the original design so as to get an optimal extended design? The main objective of the present paper is to provide an answer to this question.

1. Introduction

Uniform design has been widely used in manufacturing, system engineering, pharmaceutics and natural sciences since it was proposed by Fang (1980) and Wang and Fang (1981). The uniform experimental design is one of the most popular methodologies to construct effective designs with high-dimensional inputs and limited resources. A uniform design scatters its design points evenly on the experimental domain according to some discrepancy measure. A design $\mathscr{P}$ in a specified design space $\mathbb{P}$ is said to be a uniform design under some discrepancy if it minimizes the discrepancy among all designs in $\mathbb{P}$. There are several different discrepancies defined, among which the Lee discrepancy ($\mathcal{LD}$) (Zhou, Ning, & Song, 2008) and wrap-around L_2-discrepancy ($\mathcal{WD}$) (Hickernell, 1998) have been regarded more reasonable and practicable. These measures remain invariant under reordering of runs, relabeling coordinates and coordinate shift.

Much effort has been spent on the measurement of the design uniformity and the construction of efficient uniform designs. The problem of finding designs with minimum discrepancy is very difficult due to the computational intractability of finding a globally optimal solution, and also due to the lack of good benchmarks. To search an optimal design is an NP hard problem in the sense of computation complexity. Therefore, for reducing the computation complexity some structure of experimental points has to be considered. Instead of optimizing over a very large set of all possible n-run designs, x^n, one may find a very good design by considering a much smaller candidate set provided that it contains low discrepancy designs. One of such sets is the set of designs of balanced levels, or called U-type designs, proposed by Fang and Hickernell (1995).

* Corresponding author at: Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt.
E-mail addresses: amelsawah@uic.edu.hk, a.elsawah@zu.edu.eg, a_elsawah85@yahoo.com (A.M. Elsawah).

http://dx.doi.org/10.1016/j.jkss.2016.05.003

A symmetrical design of balanced levels $\mathcal{B}(n; q^m)$ corresponds to an $n \times m$ matrix $X = (x_1, \ldots, x_m)$ such that each column x_i takes values from a set of q integers, say $0, 1, \ldots, q-1$, equally often. Denote by $\mathbb{B}(n; q^m)$ the set of all $\mathcal{B}(n; q^m)$ designs. By mapping $f : l \to (2l+1)/(2q), l = 0, \ldots, q-1$, the n runs are transformed into n points in $C^m = [0, 1]^m$. A balanced design $\mathcal{B}(n; q^m)$ can be viewed as a design with one dimensional uniformity, that is, in each dimension, the distribution of the n points is uniform.

On the other hand, throughout our paper we give another set of designs which called a symmetrical nearly balanced design $\mathcal{N}(n; q^m)$ corresponds to an $n \times m$ matrix $X = (x_1, \ldots, x_m)$ such that each column x_i takes values from a set of q integers, say $0, 1, \ldots, q-1$, equally often as possible. Denote by $\mathbb{N}(n; q^m)$ the set of all $\mathcal{N}(n; q^m)$ designs. It is noted that any design $\mathcal{N} \in \mathbb{N}(n; q^m)$ may be balanced or nearly balanced depending upon whether n is a multiple of q or not, respectively, i.e., when n is a multiple of q the design is balanced and when n is not a multiple of q, we can find a design $\mathcal{B} \subset \mathcal{N}$ such that $\mathcal{B} \in \mathbb{B}(t; q^m)$, where t is the integral part of n/q (i.e., $\mathcal{B}$ is a balanced design with t runs).

In the last few years, there has been considerable interest in trying to explore lower bounds for different discrepancies. It is an important issue to find good lower bounds for the discrepancy measure of uniformity, because lower bounds can be used as benchmarks in searching for uniform designs. A design whose discrepancy value achieves a sharp lower bound is a uniform design with respect to this discrepancy. Many authors tried to find good lower bounds for various discrepancies. Recently, Elsawah and Qin (2014, 2015a) obtained a new lower bound of the centered L_2-discrepancy under four-level balanced designs and the first lower bound of the centered L_2-discrepancy for mixed two and three levels balanced designs respectively. Subsequently, Elsawah and Qin (2015b) obtained new lower bounds of the mixture L_2-discrepancy based on symmetric two-, three- and four-level balanced designs. Finally, Elsawah and Qin (2016b) discussed the uniformity of mixed two and three levels balanced designs in view of the mixture L_2-discrepancy.

2. Structure and possible strategies for the problem

Suppose that an experimenter begins with an experiment using a three-level uniform (nearly uniform) balanced design $\mathcal{B} \in \mathbb{B}(n; 3^m)$. After the experiment is over or during the experimentation, some additional resources become available and the experimenter can afford r more runs to the design $\mathcal{B}$. The design consisting of r runs is called as added design, denoted by $\mathcal{A} \in \mathbb{N}(r; 3^m)$. It is to be noted that the design $\mathcal{A}$ may be balanced or nearly balanced depending upon whether r is a multiple of 3 or not, respectively, i.e., the levels of each factor occur as equally often as possible. The full design obtained by augmenting the runs of the design $\mathcal{A}$ to those of the original uniform or nearly uniform design $\mathcal{B}$ is called as extended design, denoted by $\mathcal{E}$, that is, $\mathcal{E} = (\mathcal{B}'\mathcal{A}')'$. The extended design will also be balanced or nearly balanced depending upon whether $n + r$ (r) is a multiple of 3 or not, respectively. Denote by $\mathbb{E}(n + r; 3^m)$ the set of all extended designs.

An illustrative example. Suppose that an experimenter begins the experimentation using a uniform design of 5 factors with 3 levels and 9 runs. According to the available resources he/she chose a suitable design table ($\mathfrak{T}$) from the uniform design site (http://uic.edu.hk/isci/UniformDesign/UD%20Tables.html). After the experiment is over or during the experimentation, some additional resources become available and the experimenter can include 3 more runs to the design. From the uniform design site, the following design table ($\mathfrak{B}$) is the uniform design for 12 runs. We can see that, the 12 runs are completely different than the 9 runs, i.e., the level combinations are completely different, where there are just one common run (1 0 1 2 0). This means that there is no way to add the new 3 runs to his/her original 9 runs uniform design ($\mathfrak{T}$) so as to get the existing 12 runs uniform design ($\mathfrak{B}$). Therefore, there is a need to obtain a new approach for constructing uniform or nearly uniform designs obtained by adding new runs to an existing uniform design.

$$\mathfrak{T} = \begin{bmatrix} 1 & 1 & 0 & \underline{\mathbf{1}} & 2 & 0 & 2 & 2 & 0 \\ 2 & 2 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 2 & 1 \\ 1 & 0 & 0 & \underline{\mathbf{1}} & 2 & 1 & 0 & 2 & 2 \\ 0 & 2 & 0 & \underline{\mathbf{2}} & 0 & 1 & 1 & 1 & 2 \\ 2 & 1 & 0 & \underline{\mathbf{0}} & 1 & 1 & 2 & 0 & 2 \end{bmatrix}' \implies \mathfrak{B} = \begin{bmatrix} 1 & 2 & 0 & 1 & 1 & 0 & \underline{\mathbf{1}} & 0 & 2 & 2 & 2 & 0 \\ 2 & 0 & 1 & 1 & 1 & 0 & \underline{\mathbf{0}} & 0 & 1 & 2 & 2 & 2 \\ 1 & 0 & 2 & 2 & 0 & 1 & \underline{\mathbf{1}} & 2 & 0 & 1 & 2 & 0 \\ 0 & 0 & 2 & 0 & 2 & 1 & \underline{\mathbf{2}} & 1 & 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 2 & 2 & 2 & \underline{\mathbf{0}} & 0 & 0 & 2 & 1 & 1 \end{bmatrix}'.$$

The extended designs have been applied into computer experiments, microarray experiments and numerical integration, see Durrieu and Briollais (2009), Loeppky, Moore, and Williams (2010), and Tong (2006). In particular, Ji, Alaerts, Xu, Hu, and Heyden (2006) described a sequential procedure for the method of development of fingerprints based on a uniform design approach, in which the sequential uniform design is used to reach the global optimum for a separation.

A natural question is how does the experimenter pick the additional runs and augment the original design $\mathcal{B}$ so as to get an extended design $\mathcal{E}(n + r; 3^m)$ which is uniform or nearly uniform under the given measure of uniformity? Thus, our objective is to obtain an extended uniform (nearly uniform) design $\mathcal{E}(n+r; 3^m)$ given that the original design $\mathcal{B} \in \mathbb{B}(n; 3^m)$ is a uniform or a nearly uniform design. This situation can be handled in the following three different possible approaches.

1. **The first possible approach:** We may generate a uniform (nearly uniform) design directly with $n + r$ runs using the new lower bounds in Theorem 7 (given below) depending upon $(n + r)$ is a multiple of 3 or not for evaluating the efficiency of the design $\mathcal{N}(n + r; 3^m)$.
   This approach is not a feasible solution because the experimenter has already run the experiment with n design points and r additional runs have to be added with the already used experimented n runs (cf. The illustrative example).