Journal of the Korean Statistical Society 45 (2016) 168–175

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

A new class of two-level optimal extended designs

Hong Qin [a,*], Tingxun Gou [a], Kashinath Chatterjee [b]

[a] *Faculty of Mathematics and Statistics, Central China Normal University, PR China*

[b] *Department of Statistics, Visva-Bharati University, Santiniketan, India*

ARTICLE INFO

Article history:
Received 3 March 2015
Accepted 17 September 2015
Available online 21 October 2015

AMS 2000 subject classifications:
primary 62K15
secondary 62K10

Keywords:
Centered L_2-discrepancy
Lower bound
Two-level uniform design
Wrap-around L_2-discrepancy

ABSTRACT

The purpose of this article is to introduce a new class of two-level optimal extended designs obtained by adding few runs to an existing two-level uniform design. The extended design is a union of two designs belonging to different classes. New lower bounds to the centered and wrap-around L_2-discrepancies of extended designs are obtained. Some examples for optimal extended designs are also included.

1. Introduction

Simulation based on computer technology has been widely used in engineering and high-tech development. Design and modeling of computer experiments have been paid much attention in the literature, see Bates, Buck, Riccomagno, and Wynn (1996), Fang, Li, and Sudjianto (2006) and Koehler and Owen (1996) for a comprehensive review. Computer models are often used to describe complicated physical phenomena encountered in science and engineering. The uniform design seeks experimental points to be uniformly scattered in the experimental domain. It is widely accepted especially in situations where little knowledge is known about the function to be modeled. Its practical success is due to its economical and flexible experimental runs to study many factors with high levels simultaneously.

Let $\mathcal{U}_0(n, 2^m)$ be a class of two-level U-type designs with n runs and m factors. This means in any design $U \in \mathcal{U}_0(n, 2^m)$ the levels of each factor occur equally often. Similarly, let $\mathcal{U}_1(n, 2^m)$ be a class of two-level nearly U-type designs with n runs and m factors which means in any design $U \in \mathcal{U}_1(n, 2^m)$ the levels of each factor occur as equally often as possible, where "as equally often as possible" means that the difference between the times of two levels respectively occurred is at most 1. A U-type (nearly U-type) design d is called an optimal (or uniform) design under a given measure of uniformity provided that it has the best uniformity over $\mathcal{U}_u(n, 2^m)$, $u = 0, 1$. For a ready reference, mention may be made to Fang (1980), Fang and Hickernell (1995), Fang, Lin, Winker, and Zhang (2000), Fang and Mukerjee (2000) and Winker and Fang (1998).

Suppose that an experimenter begins the experimentation using a U-type design $U \in \mathcal{U}_0(n, 2^m)$ which may be optimal or nearly optimal, where "nearly optimal" means that the value of a given measure of uniformity of a U-type design is extremely close to its best uniformity value. Also suppose that after the experiment is over or during the experimentation, some additional resources become available and the experimenter can afford to include $r(=2l + u, u = 0, 1)$, supposed to

* Corresponding author.
E-mail address: qinhong@mail.ccnu.edu.cn (H. Qin).

http://dx.doi.org/10.1016/j.jkss.2015.09.003
1226-3192/

be a small integer, more runs to the design U. Let $\mathcal{U}^E(n+r, 2^m)$ be a class of two-level $(n+r)$-run extended designs in the sense that any design $U^E \in \mathcal{U}^E(n+r, 2^m)$ is obtained by adding r runs to an optimal or nearly optimal design $U \in \mathcal{U}_0(n, 2^m)$ and the design U^E is at least nearly U-type design. This means that the design U^E is such that the levels of each factor appears in the extended design as equally often as possible.

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 the experimenter will choose the additional runs and augment the original design U so as to get an extended design $U^E \in \mathcal{U}^E(n+r, 2^m)$ which is optimal or nearly optimal under the given measure of uniformity? To give an answer to this question, we have the following options.

(a) The experimenter may choose an optimal design $U \in \mathcal{U}_u(n+r, 2^m)$, $u = 0, 1$, and run the experiment with n runs and then add r (even or odd) more runs afterward.

(b) Generate a uniform (nearly uniform) design for the added r runs $U \in \mathcal{U}_u(r, 2^m)$ using the existing lower bounds. Use this design and augment this to the original design to get $U^E \in \mathcal{U}^E(n+r, 2^m)$.

(c) Addition of r (even or odd) runs to the already used optimal U-type design so that the resulting extended design will be optimal or nearly optimal over the class $\mathcal{U}^E(n+r, 2^m)$.

Option (a) is not a feasible solution because the experimenter has already run the experiment with n design points and r, not known in advance, additional runs have to be added with the already experimented runs. Moreover, option (b) is also not a good choice in general because the resulting design U^E may not be optimal, which will be explained in Remark 3. Thus, we are left with the only option, i.e., option (c).

The main objective of the present paper is to provide answer to the question and we choose option (c) for this purpose. Here we present the following definition to make clarity of the objective of the paper.

Definition 1. A design $U^E \in \mathcal{U}^E(n+r, 2^m)$ is said to be optimal if it attains the minimum discrepancy value centered L_2-discrepancy (CD for short) and wrap-around L_2-discrepancy (WD for short)).

The present paper is organized as follows. Section 2 describes in brief the uniformity measures like the centered L_2-discrepancy and the wrap-around L_2-discrepancy and also their existing lower bounds. Section 3 provides lower bounds to the uniformity measures. Section 4 presents some examples and finally Section 5 gives a concluding remark.

2. Uniformity measures and their existing lower bounds

Let $\mathcal{D}_0(n, 2^m)$ be a class of designs with n runs and m factors in which a design corresponds to an $n \times m$ array such that each of entries in each column takes values from a set of $\{0, 1\}$ equally often. Similarly, let $\mathcal{D}_1(n, 2^m)$ be a class of designs involving m factors and n runs in which a design corresponds to an $n \times m$ array such that each of entries in each column takes values from a set of $\{0, 1\}$ as equally often as possible. A typical level combination of a design $d \in \mathcal{D}_u(n, 2^m)$, $u = 0, 1$, is represented as $x = (x_1, x_2, \ldots, x_m)$, where $x_j \in \{0, 1\}$, $1 \le j \le m$. Let V be the set of all 2^m $(=v)$ level combinations (runs) in the lexicographic order. A two-level U-type design $U \in \mathcal{U}_u(n, 2^m)$, $u = 0, 1$, say, $U(n, 2^m)$ corresponds to a design $d \in \mathcal{D}_u(n, 2^m)$ such that $U(n, 2^m)$ is an $n \times m$ array with entries from the set $\{1/4, 3/4\}$ and in each column each entry appears at least as equally often as possible. It is to be noted that d and $U(n, 2^m)$ can be mapped through $u_j = (2x_j + 1)/4$, $1 \le j \le m$. A U-type design $U(n, 2^m)$ can be viewed as a design with one dimensional uniformity, that is, in each dimension, the distribution of the n points is uniform.

For any $x \in V$ and $d \in \mathcal{D}_u(n, 2^m)$, $u = 0, 1$, let $n_d(x)$ be the number of times that the level combination x occurs in d and n_d be the $v \times 1$ vector with elements $n_d(x)$ arranged in the lexicographic order. Let Z be the design matrix corresponding to a design $d \in \mathcal{D}_u(n, 2^m)$, $u = 0, 1$ with entries ± 1. It is to be noted that $Z'1_n = \pm u 1_m$, where 1_n is an $n \times 1$ vector with all elements unity. For any design $d \in \mathcal{D}_u(n, 2^m)$ and for $1 \le i, j \le n$, let c_{ij} be the number of places where the entries of the ith and the jth rows of d coincide. Then, it is easy to observe that, for $1 \le i \le n$, $c_{ii} = m$.

For a design $d \in \mathcal{D}_u(n, 2^m)$ or equivalently for any $U \in \mathcal{U}_u(n, 2^m)$, its centered and wrap-around L_2-discrepancy values, denoted as $CD_2(d)$ and $WD_2(d)$, can respectively be expressed in the following closed forms

$$[CD_2(d)]^2 = \left(\frac{13}{12}\right)^m - \frac{2}{n}\sum_{i=1}^{n}\prod_{l=1}^{m}\left(1 + \frac{1}{2}\left|u_{il} - \frac{1}{2}\right| - \frac{1}{2}\left|u_{il} - \frac{1}{2}\right|^2\right)$$
$$+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\prod_{l=1}^{m}\left(1 + \frac{1}{2}\left|u_{il} - \frac{1}{2}\right| + \frac{1}{2}\left|u_{jl} - \frac{1}{2}\right| - \frac{1}{2}\left|u_{il} - u_{jl}\right|\right)$$

and

$$[WD_2(d)]^2 = -\left(\frac{4}{3}\right)^m + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\prod_{l=1}^{m}\left(\frac{3}{2} - |u_{il} - u_{jl}|(1 - |u_{il} - u_{jl}|)\right),$$

where for $1 \le i \le n$ and $1 \le l \le m$, $u_{il} = (2x_{il} + 1)/4$ and $u_{jl} = (2x_{jl} + 1)/4$.