STAPRO: 8063 Model 3G pp. 1–7 (col. fig: NIL)

ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A general construction for nested Latin hypercube designs[☆]

Jin Xu, Xiaojun Duan [*], Zhengming Wang, Liang Yan

College of Science, National University of Defense Technology, Changsha, 410073, Hunan, China

ARTICLE INFO

Article history:
Received 4 July 2017
Received in revised form 29 October 2017
Accepted 31 October 2017
Available online xxxx

Keywords:
Computer experiments
Latin hypercube designs
Nested designs
Sampling property

ABSTRACT

We propose a new construction for nested designs, called General Nested Latin Hypercube designs (GNLHs). Such designs contain nested Latin hypercube designs as special cases. Besides achieving maximum uniformity in one dimension, each layer of GNLHs is flexible in run sizes. Moreover, theoretical results and numerical simulations show that GNLHs perform well on the sampling variance.

1. Introduction

Computer experiments are playing a significant role in science and engineering (Fang et al., 2005; Sacks et al., 1989; Santner et al., 2013). A main goal in computer experiments is to estimate the expected output of a computer model given a distribution of inputs. *Latin hypercube designs* (LHDs) proposed in McKay et al. (1979) are feasible to address this issue. An $n \times m$ matrix $H = \left(h_{ij}\right)$ is called a Latin hypercube if each column of H is a permutation of $\{1, \ldots, n\}$ and all the columns are generated independently. The Latin hypercube design $D = \left(d_{ij}\right)$ of n runs in d factors is obtained through

$$d_{ij} = \left(h_{ij} - \epsilon_{ij}\right)/n, \text{ for } i = 1, \ldots, n, j = 1, \ldots, d \tag{1}$$

where ϵ_{ij} are independent uniform variables on $[0, 1)$. The design D achieves maximum uniformity in univariate margins: when D is projected onto each of d variables, one and only one of the n design points fall within each of the n small intervals defined by $[0, 1/n), [1/n, 2/n), \ldots, [(n-1)/n, 1)$. We refer to this design as ordinary Latin hypercube design hereinafter. Stein (1987) proved that a Latin hypercube sample provides a smaller variance for the sample mean compared with the independently identical distribution sample.

Almost all of large computer experiments can be run at different levels of accuracy: the high accuracy experiments are slow and expensive, while the low accuracy experiments are fast but less accurate. Multi-accuracy experiments have received considerable attention over the past few years, such as Dewettinck et al. (1999) for simulating a GlattGPC-1 fluidized-bed unit and Choi et al. (2008) for an aircraft design application. Many modelling methods for multi-accuracy experiments have been proposed, for example Bayesian calibration (Kennedy and O'Hagan, 2001), Bayesian hierarchical modelling (Qian and Wu, 2008) and modelling for computer and physical experiments (Reese et al., 2004). Nested Latin Hypercube Designs (NLHDs) proposed in Qian (2009) are desirable to deal with these multi-accuracy experiments because for every run in the designs of the high accuracy experiments, the responses from the low accuracy experiments are also

[☆] We thank the Co-Editor and the two referees for their valuable comments and suggestions. We thank Pro. Xu He in Chinese Academy of Sciences for his kindness in polishing our work. This paper is supported by the National Natural Science Foundation of China(No .61573367, No. 11771450)

[*] Corresponding author.
E-mail address: xj_duan@163.com (X. Duan).

https://doi.org/10.1016/j.spl.2017.10.022

available. So it is easier to refine the low accuracy experiments data with the high accuracy experiments data and use the refined data to obtain a more appropriate model. Furthermore, a nested Latin hypercube design with two layers is a special Latin hypercube design that contains a smaller Latin hypercube design as a subset, where the whole set is the first layer and the embedded smaller Latin hypercube design is the second layer. This construction guarantees that each layer of NLHDs is excellent in projective uniformity. However, NLHDs and most of their variants, such as these designs discussed in He and Qian (2011), have the restriction that the run size of the low accuracy experiments have to be a multiple of the high accuracy experiments. Kong et al. (2016) proposed a new construction for nested designs (referred to as SDs hereinafter) which accommodates arbitrary sizes for different layers, but the first layer of SDs cannot achieve maximum uniformity in univariate margins. So nested designs with more flexible run sizes and better sampling properties are required to deal with the situation that the run sizes of the low accuracy experiments are not a multiple of the run sizes of the high accuracy experiments.

In this paper, we propose general nested Latin hypercube designs which can suit for different layers of arbitrary sample sizes. We refer to these new designs as GNLHs for convenience. When the run size of the first layer is a multiple of the second layer, GNLHs are equivalent to NLHDs. Moreover, each layer of GNLHs can achieve maximum uniformity in one dimension.

The paper is organized as follows. In Section 2, we introduce the construction of GNLHs. Some statistical properties of GNLHs are shown in Section 3. Section 4 concludes this paper. All the proofs are given in Appendix.

2. Construction for GNLHs

Here are some useful definitions. For an integer $k \geq 1$, let Z_k denote the set $\{1, \ldots, k\}$. $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote the ceiling and floor function, respectively. The least common multiple and the greatest common divisor of k integers $a_i, i = 1, \ldots, k$, are denoted by $lcm\,(a_1, \ldots, a_k)$ and $gcd\,(a_1, \ldots, a_k)$, respectively. For a set S, let $|S|$ denote the number of elements of S. For a matrix A, let $A\,(:, j)$ be its jth column, $A\,(i, :)$ be its ith row, and $A\,(i, j)$ be its (i, j)th entry. For sets A and B, $A \setminus B$ denotes set A minus set B. LHD(n, d) denotes the ordinary Latin hypercube design with n runs and d factors.

For two fixed integers $n > m > 0$, let $l = lcm\,(m, n)\,, n' = l/m, m' = l/n$. We propose a nested vector $\pi_{n\times 1}$ which is composed by two vectors $\tau_{m\times 1}$ and $\rho_{(n-m)\times 1}$. Let $M_{m',n'}$ denote the $(n' - m' + 1) \times n'$ circulant matrix as follows:

$$M_{m',n'} = \begin{bmatrix} 1 & 2 & \cdots & n' \\ n' & 1 & \cdots & n'-1 \\ \vdots & \vdots & \vdots & \vdots \\ m'+1 & m'+2 & \cdots & m' \end{bmatrix}.$$

Then, the vector $\pi_{m\times 1}$ is generated in the following four steps:

Step 1: Draw a vector $v_{m\times 1}$ where $v\,(1)$ is a random integer in $Z_{n'}$, and for $i = 2, \ldots, m$ and a random integer $r \in Z_{n'-m'+1}$, $v\,(i) = M_{m',n'}(r, v\,(i-1))$.

Step 2: Draw a permutation $\tau = (\tau\,(1)\,, \ldots, \tau\,(m))^T$ on $\{v\,(i) + n'(i-1) | i = 1, \ldots, m\}$ randomly. Let $S = Z_n \setminus \{\lceil \frac{\tau(1)}{m'} \rceil, \ldots, \lceil \frac{\tau(m)}{m'} \rceil\}$.

Step 3: Draw a permutation $\rho = (\rho\,(1)\,, \ldots, \rho\,(n-m))^T$ on set $\{t\,(i) + m'\,(s_i - 1) \mid i \in Z_{n-m}\}$ where $t\,(i)$ is an arbitrary integer in $Z_{m'}$ and s_i is the ith element of S in ascending order.

Step 4: Stack τ and ρ to generate a vector $\pi_{n\times 1}$, that is: $\pi = \left(\tau^T, \rho^T\right)^T$.

For the obtained vectors, a proposition is constructed as follows:

Proposition 2.1. *Consider the vectors τ, π constructed above and $a, b \in Z_l$, we have*

1. *for any $i = 1, \ldots, m, j = 1, \ldots, m,$ and $i \neq j$*

$$\lceil \frac{\tau\,(i)}{m'} \rceil \neq \lceil \frac{\tau\,(j)}{m'} \rceil; \qquad (2)$$

2. *the probability mass function for $\pi\,(i), i = 1, \ldots, n$, is*

$$P\{\pi\,(i) = a\} = \frac{1}{l},\ \text{for all } a \in Z_l. \qquad (3)$$

Eq. (2) indicates that $|S| = n - m$, so $|\rho| = n - m$ is satisfied, which guarantees the validity of Step 3. We construct a general nested Latin hypercube of n runs in d factors by taking their columns to be d independently generated π. For three integers $n > m > 0, d > 0$, GNLH(m, n, d) denotes a general nested Latin hypercube design where the first layer has n runs and the second layer has m runs. When n is a multiple of m, $d > 0$, M becomes a $\frac{n}{m} \times \frac{n}{m}$ matrix, and simple analysis shows that GNLH(m, n, d) has the same construction as NLHD(m, n, d). An example is given to illustrate the construction. Note that each dimension of the design is generated independently through the same processes, we only construct one-dimensional GNLHs for illustration.

Please cite this article in press as: Xu J., et al., A general construction for nested Latin hypercube designs. Statistics and Probability Letters (2017), https://doi.org/10.1016/j.spl.2017.10.022.