Statistics and Probability Letters 94 (2014) 239–247

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Space-filling Latin hypercube designs based on randomization restrictions in factorial experiments

Pritam Ranjan *, Neil Spencer

Department of Mathematics and Statistics, Acadia University, Wolfville, Canada, B4P2R6

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 16 July 2014
Accepted 19 July 2014
Available online 6 August 2014

Keywords:
Computer experiments
Nearly orthogonal arrays
Spreads
Stars

ABSTRACT

Latin hypercube designs (LHDs) with space-filling properties are widely used for emulating computer simulators. Over the last three decades, a wide spectrum of LHDs have been proposed with space-filling criteria like minimum correlation among factors, maximin interpoint distance, and orthogonality among the factors via orthogonal arrays (OAs). Projective geometric structures like spreads, covers and stars of $PG(p-1, q)$ can be used to characterize the randomization restriction of multistage factorial experiments. These geometric structures can also be used for constructing OAs and nearly OAs (NOAs). In this paper, we present a new class of space-filling LHDs based on NOAs derived from stars of $PG(p-1, 2)$.

1. Introduction

Latin hypercube sampling is a statistical method for generating a collections of points from a multi-dimensional distribution, which was first proposed by McKay et al. (1979) as an alternative to random sampling in the Monte Carlo methods for numerically integrating complex multi-dimensional functions. Later on, the Latin hypercube designs (LHDs) became very popular in computer experiments for building statistical metamodels (Santner et al., 2003). Random LHDs can easily be constructed; however, not all are suitable from a modeling viewpoint, for example, if all points are aligned along the main diagonal of the input space (see Section 2.1 for details).

Since replicate runs of a deterministic computer simulator generate identical outputs, it is preferred that the design points (i.e., the set of input locations for running the simulator) are spread out to fill the input space as evenly as possible. Such a design is referred to as a *space-filling* design. In this paper, we discuss space-filling LHDs, a popular class of designs in computer experiments (see Santner et al., 2003; Fang et al., 2006; and Rasmussen and Williams, 2006 for an overview). Over the last three decades, a wide spectrum of LHDs have been proposed with different space-filling criteria, for instance, minimum correlation among factors (Iman and Conover, 1982), maximin interpoint distance (Morris and Mitchell, 1995), and orthogonality among the factors via OAs (Owen, 1992; Tang, 1993). Definition 1 of Section 2.1 formalizes the definition of an OA. The construction of LHDs with space-filling criteria like maximin distance or minimum correlation often requires computationally intensive search, whereas, OA-based LHDs are easy to construct as long as the OAs exist.

The existence of a desired OA is not always guaranteed, and the construction can also be challenging (Hedayat et al., 1999). OAs can be constructed using a variety of combinatorial objects like linear codes, difference schemes and mutually orthogonal Latin squares. Rains et al. (2002) discussed the existence and construction of OAs using a spread of a finite projective space, $\mathcal{P} = PG(p-1, q)$, and called them *geometric OAs*. The finite projective space $\mathcal{P} = PG(p-1, q)$ is the set of all p-dimensional pencils over $GF(q)$, or equivalently, a geometry whose {points, lines, planes, …, hyperplanes} are the

* Corresponding author.
E-mail address: pritam.ranjan@acadiau.ca (P. Ranjan).

http://dx.doi.org/10.1016/j.spl.2014.07.032

subspaces of V_q^p of rank $\{1, 2, 3, \ldots, p-1\}$, where V_q^p is a vector space of rank p over $GF(q)$, and the dimension of a subspace (or flat) of $\mathcal{P}$ is one less than the rank of a subspace of V_q^p. This paper focuses on a class of LHDs that are based on projective space over $GF(2)$, i.e., $PG(p-1, 2)$.

Ranjan et al. (2009) established an equivalence between 2^p factorial experiments with multiple randomization restrictions and various geometric structures of $PG(p-1, 2)$ (e.g., spreads and covers). Here, a point (or pencil) in $PG(p-1, 2)$ corresponds to a factorial effect, and a *spread* of $\mathcal{P}$ is a set of disjoint flats of $\mathcal{P}$ that covers all points of $\mathcal{P}$. For example, in a 2^4 factorial experiment, $\mathcal{P} = \{A, B, AB, C, AC, \ldots, ABCD\}$ is a $PG(3, 2)$, and $\psi = \{S_1 = \{D, BC, BCD\}, S_2 = \{C, AB, ABC\}, S_3 = \{B, ACD, ABCD\}, S_4 = \{A, BD, ABD\}, S_5 = \{CD, AC, AD\}\}$ is a spread of 1-flats of $\mathcal{P}$. Randomization restrictions at stage i of a factorial experiment is characterized by a *randomization defining contrast subspace* (RDCSS) obtained by spanning $t_i(\leq p)$ linearly independent randomization factors (or factorial effects), which is equivalent to a (t_i-1)-flat of $\mathcal{P}$ (e.g., S_i's in ψ). Such RDCSSs are similar to block defining contrast subgroups in a blocked factorial design, but have to be separate for every stage. See Section 2.2 for a detailed discussion on RDCSSs.

For efficient analysis of a multistage factorial experiment, it is desirable to construct disjoint RDCSSs. However, in many practical situations (e.g., the plutonium alloy experiment of Bingham et al., 2008), overlap among the RDCSSs cannot be avoided. For such cases, Ranjan et al. (2010) proposed designs based on a new geometric structure called a *star*—a set of distinct flats of $PG(p-1, q)$ that share a common overlap (the nucleus). A star that is also a cover (referred to as a *covering star*) of $PG(p-1, q)$ simplifies to a spread if the nucleus is empty. For example, in a 2^5 factorial experiment, $\mathcal{P} = \{A, B, AB, C, AC, \ldots, ABCDE\}$ is a $PG(4, 2)$, and $\Omega = \{R_1, R_2, R_3, R_4, R_5\}$ is a covering star with five rays, $\{R_1 = \langle D, BC, ABCDE\rangle,\ R_2 = \langle C, AB, ABCDE\rangle,\ R_3 = \langle B, ACD, ABCDE\rangle,\ R_4 = \langle A, BD, ABCDE\rangle$ and $R_5 = \langle CD, AC, ABCDE\rangle\}$, and nucleus $\pi = \{ABCDE\}$ of $\mathcal{P}$, where $\langle F_1, \ldots, F_n\rangle$ denotes the span of $F_1, \ldots, F_n$.

We have discovered a new class of space-filling LHDs that can be constructed using stars of $PG(p-1, 2)$. It turns out that a star with non empty nucleus generates near-OAs (NOAs). In general, a near-OA is an array in which the orthogonality requirement is nearly satisfied (for details, see Taguchi, 1959; Wang and Wu, 1992; Nguyen, 1996; Wu and Hamada, 2000; and Xu, 2002). In the spirit of Rains et al. (2002), we sometimes refer to these star-based NOAs as *geometric NOAs*. By following Tang's OA-based LHD construction algorithm, we construct a class of geometric NOA-based LHDs. Although such LHDs are not always very space-filling, a near orthogonality (e.g., Xu and Wu, 2001) or space-filling criterion can be used to search for a good one. To avoid the search, we also propose a set of guidelines for carefully distributing the factorial effects among RDCSSs of the star which ensures space-filling LHDs. It is worth noting that the existence of OA-based LHDs are limited to only few $n \times d$ combinations, whereas, the existence conditions for stars are less stringent.

The remainder of the paper is organized as follows. Section 2 presents an overview of LHDs, RDCSSs in a 2^p factorial experiment, and spreads and stars of a $PG(p-1, q)$. In Section 3, we establish theoretical results for the existence and an algorithm of the construction of geometric-NOAs. Section 4 concludes the paper with a few remarks.

2. Background

This section starts with a brief review on random LHDs and OA-based LHDs. Then a few results are presented to establish the equivalence between a multistage factorial design with randomization restrictions and geometric structures of $PG(p-1, 2)$.

2.1. Latin hypercube designs

Let $L(n, d)$ be an LHD with n runs in d factors (dimension of input space), where L_{ij} denotes the level of factor j in the ith experimental run, and each factor includes n uniformly spaced levels. In computer experiments, the input spaces are typically bounded hyper-rectangles, and can be transformed to unit hypercubes. A random $L(n, d)$ in $[0, 1]^d$ has $L_{ij} = (\pi_j(i) - u_{ij})/n$ for $1 \leq j \leq d$ and $1 \leq i \leq n$, where $u_{ij} \sim \text{Unif}(0, 1)$ and $(\pi_j(1), \ldots, \pi_j(n))$ is a random permutation of $\{1, \ldots, n\}$ (see Tang, 1993 for details). Ignoring the Unif(0, 1) perturbations, there are $(n!)^d$ distinct LHDs.

Although LHDs have a nice one-dimensional projection property, that is, one point each in $((i-1)/n, i/n)$ for $1 \leq i \leq n$, random LHDs can be quite undesirable from a modeling viewpoint. Fig. 1 presents two realizations of random LHDs in $[0, 1]^2$. The points in Fig. 1(a) are distributed throughout the whole space (space-filling), but the points in Fig. 1(b) are concentrated along the main diagonal.

In this paper, we propose a class of LHDs, called star-based LHDs, which are space-filling under certain conditions. Star-based LHDs are generalizations of the OA-based LHDs.

Definition 1. An $n \times d$ array $\mathcal{A}$ denoted by $\text{OA}(n, s_1 s_2 \cdots s_d, r)$ is said to be a strength r OA with n runs and d factors, if factor j has s_j levels $\{0, \ldots, s_j - 1\}$ and each $n \times r$ subarray contains every possible r-tuple an equal number of times.

The special case of $s_1 = s_2 = \cdots = s_d$ corresponds to a symmetric OA denoted by $\text{OA}(n, s, d, r)$. Although mixed-level (or asymmetric) OAs have been investigated in recent years (Hedayat et al., 1999), it is a less explored area than symmetric OAs.

A simple existence condition of an asymmetric OA of strength r follows from the strength aspect of Definition 1, that is, n must be a multiple of $s_1^{x_1} s_2^{x_2} \cdots s_d^{x_d}$ for every set of $x_1, \ldots, x_d \in \{0, 1\}$ such that $\sum_{j=1}^{d} x_j \leq r$. Another popular existence