

OA-based Latin Hypercubes

Boxin Tang

**IIQP Research Report
RR-91-10**

November 1991

OA-BASED LATIN HYPERCUBES

Boxin Tang

Department of Statistics and Actuarial Science
University of Waterloo

ABSTRACT

The growing popularity of latin hypercubes during the last decade both in simulations and in experimental designs is largely due to their attractive properties in one dimension. Their possibly anomalous behavior in multi-dimensions also causes some concerns among statisticians. Toward this a remedial measure has been proposed by several authors that aims at controlling the correlations in latin hypercubes. The problems with this approach are that there is no direct connection between uniformity and low correlation and that it works only for two-dimensions. In this paper we introduce orthogonality in latin hypercubes by exploiting the use of orthogonal arrays. The resulting designs and sampling scheme are referred to as U-designs and U-sampling respectively. Our approach is preferable in many ways. We show that U-sampling when used in Monte Carlo integration offers a substantial improvement over latin hypercube sampling. Many examples of U-designs are presented. A procedure for generating U-samples is also provided.

Key words and phrases: latin hypercubes, orthogonal arrays, OA-based latin hypercubes, U-designs, U-sampling.

1 Introduction

To motivate our approach, we begin by describing and reviewing briefly some developments in two related statistical areas.

1.1 Simulation

In many scientific and technological fields, we are often confronted with the problem of evaluating a complex integral over a multi-dimensional domain. Among a whole collection of numerical integration techniques, the Monte Carlo method is especially useful and often competitive for problems of high dimensions(Davis and Rabinowitz, 1984 chapter 5.10). The problem is formulated as follows:

Consider a deterministic function $Y = f(X)$ where $Y \in R$, $X \in R^m$ and f is known but expensive to compute. The random vector X has m statistically independent components X^1, \dots, X^m and X^j has a known distribution F_j , $j = 1, \dots, m$. Let $F = \prod_{j=1}^m F_j$ denote the distribution of X . We want to estimate the mean of Y , $\mu = E(Y)$.

The simplest way by Monte Carlo is to draw X_1, \dots, X_n independently from F and use $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ as the estimate of μ , where $Y_i = f(X_i)$. Mckay, Conover and Beckman(1979) introduce latin hypercube sampling(LHS) as an alternative to i.i.d. sampling. They show that LHS can result in the variance reduction of \bar{Y} when f is monotone in each variable. Stein(1987) obtains a more informative result. If we define the main effects as $f_j(X) = E(f(X)|X^j) - \mu$ and write

$$f(X) = \mu + \sum_{j=1}^m f_j(X) + r(X),$$

where

$$r(X) = f(X) - \mu - \sum_{j=1}^m f_j(X).$$

Stein shows that the variance of \bar{Y} under LHS is $n^{-1}\text{Var}[r(X)] + o(n^{-1})$ that is asymptotically smaller than the i.i.d. variance $n^{-1}\text{Var}[f(X)]$.

The main feature of LHS is that it stratifies each dimension simultaneously. Stein's result simply states that by using LHS the main effects are filtered out. One may expect that if stratification is also achieved on each two-dimension, one may filter out another part from the error term. Owen(1990) demonstrates that some bilinear terms can be filtered out by matching the sample and the population correlations. Our approach aims directly at stratification for multi-dimensions and hence a more general result can be obtained.

1.2 Experimental Design

In the theory and practice of experimental design, it is well-known that when the assumed model is suspect, we are led to concentrating on and minimizing the bias part of the mean square error. This usually can be achieved by scattering design points uniformly in the design region. See Box and Draper(1959) and Sacks and Ylvisaker(1984) for references. Orthogonal array designs are extensively used for planning experiments in industry and the successes achieved by using them are at least in part due to their uniformity properties. However, when a large number of factors are involved in an experiment and only a few of them are important, orthogonal array designs when projected onto the subspace spanned by the important factors might result in replication of points. This is undesirable when the bias is more serious than the variance, and is a disaster for deterministic computer experiments. In this case, latin hypercube designs are the preferred alternatives to adopt(Welch, Buck,

Sacks, Wynn, Mitchell, and Morris, 1992). However, the projections of design points onto even two-dimensions cannot be guaranteed to be uniformly scattered. Current methods for improving latin hypercube designs control the correlations, see Iman and Conover(1982) and Owen(1990). Two objections to this approach are that there is no direct connection between uniformity and low correlation and that it works solely for two-dimensions.

We introduce orthogonality in latin hypercubes by exploiting the use of orthogonal arrays. Our approach seems to be preferable to the correlation approach in many ways. It not only eliminates the drawbacks of the correlation approach but also has some important statistical properties. We will discuss these problems in later sections.

By using these OA-based latin hypercubes, the resulting designs and sampling scheme are referred to as U-designs and U-sampling, respectively. We introduce the basic concepts of U-designs and U-sampling in Section 2 present various examples in Section 3. Some theoretical results on U-sampling are provided in Section 4. Finally in Section 5 we give generalization and discussion of our approach.

2 Basics of U-designs and U-sampling

2.1 Mathematical Preliminaries

An $n \times m$ matrix A , with entries from a set of $s \geq 2$ elements, is called an orthogonal array with strength r , size n , m constraints and s levels if each $n \times r$ submatrix of A contains all possible $1 \times r$ row vectors with the same frequency λ . The number λ is called the index of the array. Clearly $n = \lambda s^r$. The array is denoted by $OA(n, m, s, r)$.

Denote the s elements by $1, 2, \dots, s$. Let P_n be the set of permutations of n objects

$\{1, \dots, n\}$. A generic element of P_n will be denoted by ρ with or without a subscript, ρ permuting $\{1, \dots, n\}$ to $\{\rho(1), \dots, \rho(n)\}$. Let Γ_m be the set of all sequences with entries from P_n of length m . Since $n = \lambda s^r$, we can define a mapping Z from $\{1, \dots, n\}$ onto $\{1, \dots, s\}$ as follows:

$$Z(i) = \begin{cases} 1 & i = 1, 2, \dots, q, \\ 2 & i = q + 1, q + 2, \dots, 2q, \\ \vdots & \vdots \\ s & i = (s - 1)q + 1, (s - 1)q + 2, \dots, n, \end{cases}$$

where $q = n/s$.

An element $(\rho_1, \rho_2, \dots, \rho_m)$ in Γ_m is said to be an orthogonal sequence with strength r , if the matrix

$$\begin{pmatrix} Z(\rho_1(1)) & Z(\rho_2(1)) & \cdots & Z(\rho_m(1)) \\ Z(\rho_1(2)) & Z(\rho_2(2)) & \cdots & Z(\rho_m(2)) \\ \vdots & \vdots & & \vdots \\ Z(\rho_1(n)) & Z(\rho_2(n)) & \cdots & Z(\rho_m(n)) \end{pmatrix}$$

forms an $OA(n, m, s, r)$.

For each $2 \leq r \leq m$, we define

$$\Gamma_m^r = \{\rho \in \Gamma_m \mid \rho \text{ is an orthogonal sequence with strength } r\}.$$

Clearly, $\Gamma_m^m \subseteq \Gamma_m^{m-1} \subseteq \cdots \subseteq \Gamma_m^2 \subseteq \Gamma_m$.

2.2 The Concepts of U-designs and U-sampling

Now, we can define a latin hypercube design(LHD) as follows: for any sequence (ρ_1, \dots, ρ_m) in Γ_m , we say the matrix

$$\begin{pmatrix} \rho_1(1) & \rho_2(1) & \cdots & \rho_m(1) \\ \rho_1(2) & \rho_2(2) & \cdots & \rho_m(2) \\ \vdots & \vdots & & \vdots \\ \rho_1(n) & \rho_2(n) & \cdots & \rho_m(n) \end{pmatrix} \quad (1)$$

is an LHD.

If we allow ρ_1, \dots, ρ_m to be chosen freely, the resulting LHD could be very undesirable. For example, the situation in Figure 1 of Section 3 could arise, where $n = 4$, $m = 2$, and $\rho_1 = \rho_2$.

To rectify this problem, we introduce

Definition 1 For any sequence (ρ_1, \dots, ρ_m) in Γ_m^r , we say the matrix in (1) is a U-design with parameters (n, m, s, r) .

We avoid giving the same names for (n, m, s, r) as we do for an orthogonal array, because they do not all have an obvious statistical interpretation. The most notable is s which does not refer to the number of levels of the design(actually the design has n levels), although the others n, m, r have similar interpretations.

From the definition, a U-design is a special LHD that forms an orthogonal array design after grouping n original levels into n/s levels in a systematic way. One obvious implication is that a U-design can achieve some kind of uniformity in each r -dimension. Since U-designs also retain some sort of balance properties inherent from the parent orthogonal array, it is possible to use the grouped levels to perform a preliminary data analysis to identify the

important factors. After the important factors have been identified, one can revert to the original levels to model the response by a regression or Bayesian model.

To motivate the definition of U-sampling, we first describe LHS as follows: for each $j = 1, \dots, m$, draw independently $X_i^j \sim U((i-1)/n, i/n)$, $i = 1, \dots, n$, such that all X_i^j are independent. Next a sample of size m is drawn from P_n by simple random sampling with replacement; the resulting permutation sequence is denoted by (ρ_1, \dots, ρ_m) . Then

$$X_i = (X_{\rho_1(i)}^1, X_{\rho_2(i)}^2, \dots, X_{\rho_m(i)}^m)$$

$i = 1, \dots, n$, is a desired latin hypercube sample.

Definition 2 *Instead of drawing a permutation sequence randomly from Γ_m , we draw a permutation sequence randomly from Γ_m^r , and form the n inputs in the same way. Then the sample so obtained is called a U-sample and the corresponding sampling scheme is called U-sampling with strength r .*

The essential difference between the U-sampling and LHS is that U-sampling excludes those permutation sequences not contained in Γ_m^r . This is in line with controlled sampling in sample survey theory. For a recent reference on controlled sampling, see Rao and Nigam(1990). Intuitively, the permutation sequences in Γ_m^r are preferred since some kind of stratification is achieved for all r dimensions. We will give some theoretical justifications in Section 4.

To estimate the mean of Y , μ , we still use the sample mean, that is,

$$\hat{\mu} = \frac{1}{n}(Y_1 + \dots + Y_n),$$

where $Y_i = f(X_i)$, $i = 1, \dots, n$.

Before going further, we present some examples in Section 3 for illustration. A method for obtaining all possible U-designs by starting with an orthogonal array is also described through these examples.

3 Examples

Example 1 Consider the case of two factors and four runs, that is, $n = 4$ and $m = 2$. We can take $s = 2$ and the mapping Z as follows:

$$Z(i) = \begin{cases} 1 & i = 1, 2, \\ 2 & i = 3, 4. \end{cases}$$

Since

$$\begin{pmatrix} Z(1) & Z(2) & Z(3) & Z(4) \\ Z(1) & Z(3) & Z(2) & Z(4) \end{pmatrix}^t = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 2 \end{pmatrix}^t$$

is an OA(4,2,2,2), the following matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}^t$$

is a U-design. And since

$$\begin{pmatrix} Z(1) & Z(2) & Z(3) & Z(4) \\ Z(1) & Z(2) & Z(4) & Z(3) \end{pmatrix}^t = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{pmatrix}^t$$

does not form an orthogonal array, the matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}^t$$

is not a U-design.

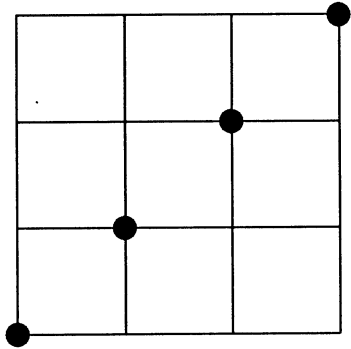


Figure 1

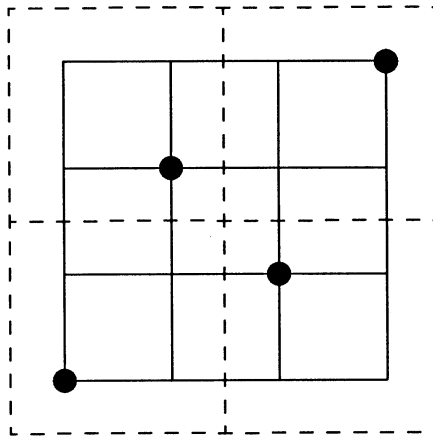


Figure 2

These two designs are represented by Figures 2 and 3. Figure 2 represents a U-design and Figure 3 an LHD which is not a U-design. On intuitive grounds, the former is preferred to the latter. Indeed suppose that the area within the dashed-line-box is the design region. The design points in Figure 2 are more uniformly scattered than those in Figure 3 in the two-dimensional region in the sense that each small dashed-line-box contains only one design point.

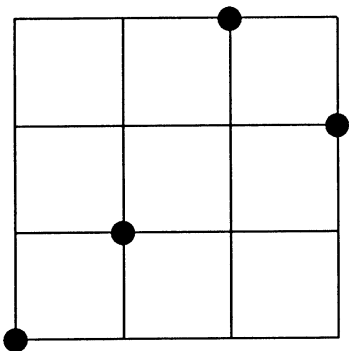


Figure 3

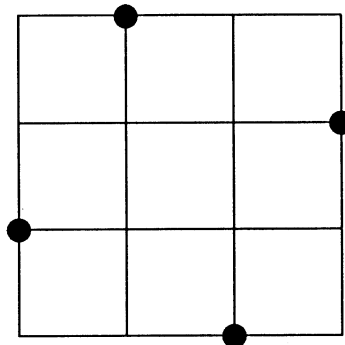


Figure 4

Another U-design, more attractive in the sense that the design points are more uniformly scattered than those given in Figure 2, is

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}^t,$$

which is represented by Figure 4.

Example 2 Let $n = 9$, $m = 4$. We choose $s = 3$ and

$$Z(i) = \begin{cases} 1 & i = 1, 2, 3, \\ 2 & i = 4, 5, 6, \\ 3 & i = 7, 8, 9. \end{cases}$$

It is easy to check that the matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 6 & 9 & 2 & 5 & 8 & 1 & 4 & 7 \\ 1 & 4 & 7 & 5 & 8 & 2 & 9 & 3 & 6 \\ 1 & 7 & 4 & 5 & 2 & 8 & 9 & 6 & 3 \end{pmatrix}^t$$

is a U-design with parameters $(9,4,3,2)$. The projections of points onto the first two columns of the matrix can be represented by Figure 5.

Note that for each column of the $OA(9,4,3,2)$, if we replace three 1's by any permutation of $\{1,2,3\}$, replace three 2's by any permutation of $\{4,5,6\}$, and replace three 3's by any permutation of $\{7,8,9\}$, the resulting matrix is a U-design with parameters $(9,4,3,2)$. This gives a simple procedure for constructing all possible U-designs by starting with an orthogonal array.

Example 3 Let $n = 8$, $m = 3$. We choose $s = 2$ and

$$Z(i) = \begin{cases} 1 & i = 1, 2, 3, 4, \\ 2 & i = 5, 6, 7, 8. \end{cases}$$

It is easy to check that the matrix

$$\begin{pmatrix} 4 & 3 & 2 & 1 & 8 & 7 & 6 & 5 \\ 3 & 1 & 7 & 5 & 4 & 2 & 8 & 6 \\ 1 & 5 & 2 & 6 & 3 & 7 & 4 & 8 \end{pmatrix}^t$$

is a U-design with parameters $(8,3,2,3)$.

The projections of points onto the first two columns of the matrix can be represented by Figure 6.

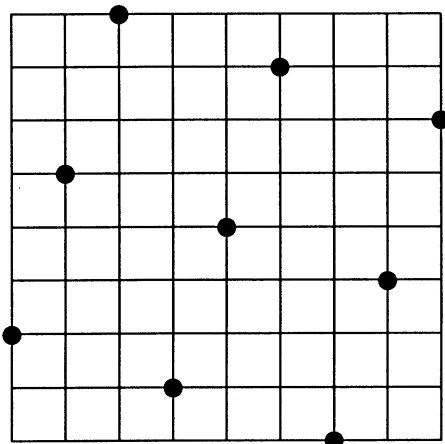


Figure 5

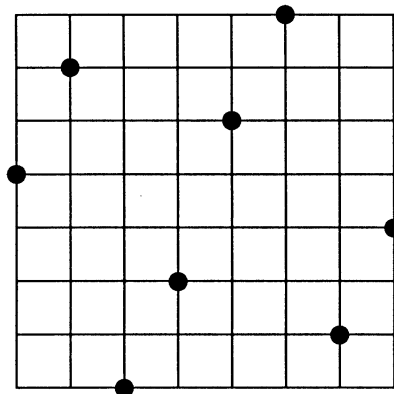


Figure 6

Example 4 Let $n = 16$, $m = 2$. We can choose s to be either 4 or 2, since both $OA(16,2,4,2)$ and $OA(16, 2, 2, 2)$ can be used to construct U-designs. It is felt that the case of $s = 4$ is preferred and recommended. Generally, larger values of s provide more balance.

Example 5 We consider a more realistic situation where $n = 49$, $m = 8$, $s = 7$, and the mapping Z is defined to be

$$Z(i) = \begin{cases} 1 & i = 1, 2, 3, 4, 5, 6, 7, \\ 2 & i = 8, 9, 10, 11, 12, 13, 14, \\ 3 & i = 15, 16, 17, 18, 19, 20, 21, \\ 4 & i = 22, 23, 24, 25, 26, 27, 28, \\ 5 & i = 29, 30, 31, 32, 33, 34, 35, \\ 6 & i = 36, 37, 38, 39, 40, 41, 42, \\ 7 & i = 43, 44, 45, 46, 47, 48, 49. \end{cases}$$

A random U-design with parameters $(49, 8, 7, 2)$ is generated using the method given in Example 2. The pairwise plots of all the two-dimensions are provided in Appendix B. In order to make a comparison, we also generate a random LHD. It is seen that the design points in the former look more uniform than those in the latter.

Example 6 This example is devoted to U-sampling with $n = 4$, $m = 2$, $s = 2$ and Z as in Example 1. The resulting U-sample is plotted in Figure 7.

In addition to stratifying each dimension, it also stratifies the two-dimensional region as shown in Figure 7.

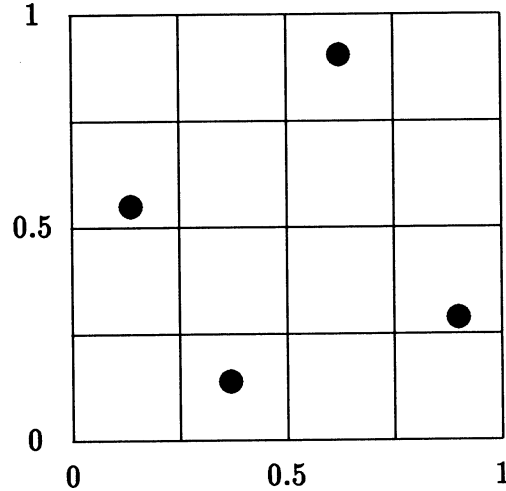


Figure 7

Before we conclude this section, we point out that construction of a U-design depends on the existence of a corresponding orthogonal array. The same remark is true of U-sampling. For related results on orthogonal arrays, we refer to Plackett and Burman(1946), Rao(1947), Bose and Bush(1952), Dey(1985) and de Launey(1986).

4 Some Results on U-sampling

We proceed to study some theoretical properties of U-sampling in this section. First we introduce a generic two-stage sampling scheme in Section 4.1 which includes both LHS and U-sampling as special cases. This general framework facilitates the theoretical development in the rest of the section and also offers the possibility of constructing new sampling schemes. In Section 4.2 some results on U-sampling are presented; comparisons with the work of Stein and Owen are also given. Finally we briefly describe a procedure for generating U-samples

in Section 4.3.

4.1 Two-stage Sampling

Recall that $Y = f(X)$ and $X = (X^1, \dots, X^m)$, where X^1, \dots, X^m are mutually independent and $X^j \sim F_j$, $j = 1, \dots, m$. Without loss of generality, we may assume that $X^j \sim U(0, 1]$, $j = 1, \dots, m$. Let

$$\mathcal{A} = \{(0, 1/n], (1/n, 2/n], \dots, (1 - 1/n, 1]\}$$

and

$$\mathcal{C} = \{A_1 \times \dots \times A_m \mid A_j \in \mathcal{A}, j = 1, \dots, m\}.$$

Members of \mathcal{C} are referred to as cells and thus n^m cells form a partition of $(0, 1]^m$.

To estimate the mean of Y , μ , for each cell in \mathcal{C} we would like to draw an X from the uniform distribution on that cell. Consideration of cost, however, leads us to use only a sample of cells and hence to consider the following two-stage sampling scheme:

Stage 1. Draw a sequence of cells C_1, \dots, C_n from \mathcal{C} using some sampling scheme satisfying

1. The random vector (C_1, \dots, C_n) is exchangeable,
2. The marginal distribution of C_i is uniform on \mathcal{C} , $i = 1, \dots, n$.

Stage 2. For each C_i , $i = 1, \dots, n$, X_i is drawn from the uniform distribution on C_i , the drawing being carried out independently for each C_i , $i = 1, \dots, n$.

Denote the variance of Y as σ^2 . Let $\bar{Y} = (Y_1 + \dots + Y_n)/n$ where $Y_i = f(X_i)$, $i = 1, \dots, n$.

Then we have the following obvious assertions.

Assertion 1. Under the two-stage sampling scheme, \bar{Y} is an unbiased estimator of μ , that is,

$$E(\bar{Y}) = \mu.$$

Assertion 2. Under the two-stage sampling scheme, the variance of \bar{Y} is given by

$$\text{Var}(\bar{Y}) = \frac{\sigma^2}{n} + \frac{n-1}{n} \text{Cov}(Y_1, Y_2).$$

The proofs of the assertions are straightforward. Assertion 1 follows immediately from the fact that the marginal distribution of X_i is uniform on $(0, 1]^m$ under the two-stage sampling, $i = 1, \dots, n$. The proof of Assertion 2 is based on the formula:

$$\text{Var}(\bar{Y}) = \frac{1}{n^2} \left(\sum_{i=1}^n \text{Var}(Y_i) + \sum_{i \neq i'}^n \text{Cov}(Y_i, Y_{i'}) \right),$$

and the two conditions in Stage 1.

It is easy to see that both LHS and U-sampling are special cases of the two-stage sampling scheme, where the i th cell C_i corresponds to the i th row of the matrix in (1) of Section 2.2. Therefore Assertions 1 and 2 hold for both of them.

4.2 Some Results on U-sampling

Let

$$\mathcal{B} = \{(0, 1/s], (1/s, 2/s], \dots, (1 - 1/s, 1]\}$$

and

$$\mathcal{L} = \{B_1 \times \dots \times B_m \mid B_j \in \mathcal{B}, j = 1, \dots, m\}.$$

Members of \mathcal{L} are referred to as large cells and thus s^m large cells form a partition of $(0, 1]^m$.

We shall refer to the members of \mathcal{C} introduced in Section 4.1 as small cells from now on.

Intuitively, it is required that s be large in order to derive the asymptotic variance of \bar{Y} under U-sampling. In this case, an economical way is to choose an orthogonal array with $\lambda = 1$. In this section we shall concentrate on the case $r = 2$ for two reasons. First, the sample size $n = s^2$ is held at its minimum. Second, the basic ideas should carry through for a general r although things become very complicated.

Now we are ready to give the following theorem.

Theorem 1 *Suppose f is bounded. Then under U-sampling we have*

$$\text{Cov}(f(X_1), f(X_2)) = -\frac{1}{n} \sum_j \text{Var}(E(f(X_1) | c(A_j)))$$

$$-\frac{1}{n} \sum_{i < j} \text{Var}\{E(f(X_1) | c(B_i \times B_j)) - E(f(X_1) | c(B_i)) - E(f(X_1) | c(B_j))\} + o\left(\frac{1}{n}\right),$$

where $n = s^2$, $A_1 \times \cdots \times A_m$ and $B_1 \times \cdots \times B_m$ are the small and large cells to which X_1 belongs, and $c(A_j)$, $c(B_j)$ and $c(B_i \times B_j)$ are the cylinder sets with bases A_j , B_j and $B_i \times B_j$ respectively.

The proof of Theorem 1 is given in the Appendix A.

Corollary 1 *If f is bounded continuous, then under U-sampling we have*

$$\text{Cov}[f(X_1), f(X_2)] = -\frac{1}{n} \sum_j \text{Var}[E(f(X) | X^j)]$$

$$-\frac{1}{n} \sum_{i < j} \text{Var}[E(f(X) | X^i, X^j) - E(f(X) | X^i) - E(f(X) | X^j)] + o\left(\frac{1}{n}\right),$$

where X is uniformly distributed on $(0, 1]^m$.

It follows from Theorem 1 and the continuity of f .

Theorem 2 *If f is bounded continuous, then under U-sampling we have*

$$\text{Var}(\bar{Y}) = \frac{\sigma^2}{n} - \frac{1}{n} \sum_j \text{Var}[f_j(X^j)] - \frac{1}{n} \sum_{i < j} \text{Var}[f_{ij}(X^i, X^j)] + o\left(\frac{1}{n}\right),$$

where

$$f_j(X^j) = E(f(X) | X^j) - \mu$$

is the main effect of X^j , and

$$f_{ij}(X^i, X^j) = E[f(X) | X^i, X^j] - f_i(X^i) - f_j(X^j) + \mu$$

is the second-order interaction between X^i and X^j .

It follows from the Corollary 1 and Assertion 2 in Section 4.1.

If we write

$$f(X) = \mu + \sum_j f_j(X^j) + \sum_{i < j} f_{ij}(X^i, X^j) + r(X), \quad (2)$$

it is easy to check that the terms on the right hand side of (2) are orthogonal to each other.

Therefore we have the corresponding variance decomposition

$$\sigma^2 = \sum_j \text{Var}(f_j) + \sum_{i < j} \text{Var}(f_{ij}) + \text{Var}(r(X)).$$

Then Stein's result(1987) shows that under LHS

$$\text{Var}(\bar{Y}) = \frac{\sigma^2}{n} - \frac{1}{n} \sum_j \text{Var}(f_j) + o\left(\frac{1}{n}\right),$$

and Theorem 2 states that under U-sampling

$$\begin{aligned} \text{Var}(\bar{Y}) &= \frac{\sigma^2}{n} - \frac{1}{n} \sum_j \text{Var}(f_j) - \frac{1}{n} \sum_{i < j} \text{Var}(f_{ij}) + o\left(\frac{1}{n}\right) \\ &= \frac{1}{n} \text{Var}[r(X)] + o\left(\frac{1}{n}\right). \end{aligned}$$

The advantage from using U-sampling instead of LHS is obvious. U-sampling is capable of filtering out *all the second-order interactions* as well as the main effects. This generalizes Stein's work. Owen(1990) demonstrates that bilinear terms can be filtered out by matching correlations of latin hypercube sample and population. To compare Theorem 2 with the Owen's work, further let

$$f_{ij}(X^i, X^j) = \gamma_{ij}(X^i - 0.5)(X^j - 0.5) + g_{ij}(X^i, X^j), \quad (3)$$

where

$$\gamma_{ij} = \frac{E[f_{ij}(X^i, X^j)(X^i - 0.5)(X^j - 0.5)]}{E[(X^i - 0.5)(X^j - 0.5)]^2}$$

so that the two terms on the right hand side of (3) are mutually orthogonal. Then Owen's result can be written as

$$\text{Var}(\bar{Y}) = \frac{1}{n} \text{Var}[r(X)] + \frac{1}{n} \sum_{i < j} \text{Var}(g_{ij}) + o\left(\frac{1}{n}\right).$$

It is seen that Theorem 2 is also more general than Owen's work.

Recently, Owen(1991) independently obtains a similar result as Theorem 2 by merely using randomized orthogonal arrays. To get the same accuracy for estimating the main effects, the number of runs required for orthogonal arrays may be too large.

4.3 A Method for Generating U-samples

It is difficult to use Definition 2 of Section 2 to generate U-samples since the number of orthogonal sequences of permutations is too large. We will give a method for doing this in this section. The method is illustrated by a simple example.

Suppose that $n = 9$ and $m = 3$. Here we take $s = 3$. Consider the orthogonal array OA(9, 4, 3, 2).

1. Draw three columns from four columns of the $OA(9, 4, 3, 2)$ by simple random sampling without replacement. Denote the three columns by a_1, a_2 and a_3 in drawn order.
2. For each column $a_j, j = 1, 2, 3$ replace the three objects 1, 2, 3 by a random permutation of them.
3. For each column $a_j, j = 1, 2, 3$ replace three 1's by a random permutation of $\{1, 2, 3\}$, replace three 2's by that of $\{4, 5, 6\}$, and replace three 3's by that of $\{7, 8, 9\}$.

Then the 9×3 matrix (a_1, a_2, a_3) gives an orthogonal sequence of permutations as defined in Section 2. The nine rows of the matrix above represents nine small cells. The work for selecting X's is trivial after the small cells have been obtained.

5 Generalization and Discussion

5.1 A Generalization of U-designs and U-sampling

There are some limitations in the previous definitions of U-designs and U-sampling. By utilizing asymmetric orthogonal arrays, the ideas still carry through. For recent references on asymmetric orthogonal arrays, see Rao(1973), Dey(1985), Wu(1989), Wang and Wu(1991) and Wu, Zhang and Wang(1990).

The notation $OA(n, r, s_1, \dots, s_m)$ is used to denote an orthogonal array of size n and strength r with s_j levels in the j th column, $j = 1, \dots, m$, and the s_j 's may not be distinct.

Define m mapping Z_j as follow:

$$Z_j(i) = \begin{cases} 1 & i = 1, 2, \dots, q_j \\ 2 & i = q_j + 1, q_j + 2, \dots, 2q_j \\ \vdots & \vdots \\ s_j & i = (s_j - 1)q_j + 1, (s_j - 1)q_j + 2, \dots, n, \end{cases}$$

where $q_j = n/s_j$, for $j = 1, \dots, m$.

Given a permutation sequence $(\rho_1, \rho_2, \dots, \rho_m)$, we say the matrix in (1) of Section 2 is an asymmetric U-design if the matrix

$$\begin{pmatrix} Z_1(\rho_1(1)) & Z_2(\rho_2(1)) & \cdots & Z_m(\rho_m(1)) \\ Z_1(\rho_1(2)) & Z_2(\rho_2(2)) & \cdots & Z_m(\rho_m(2)) \\ \vdots & \vdots & & \vdots \\ Z_1(\rho_1(n)) & Z_2(\rho_2(n)) & \cdots & Z_m(\rho_m(n)) \end{pmatrix}$$

forms an $OA(n, r, s_1, \dots, s_m)$

The generalization of U-sampling can be made accordingly. We shall not give the details here.

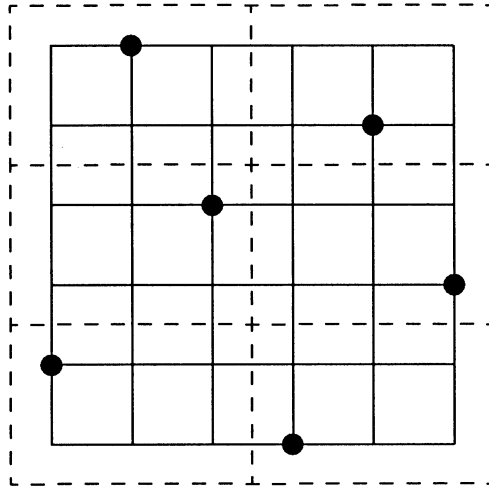
Example 7 We consider the case where $n = 6$, $m = 2$, $s_1 = 2$ and $s_2 = 3$. Two mappings are defined as follows:

$$Z_1(i) = \begin{cases} 1 & i = 1, 2, 3, \\ 2 & i = 4, 5, 6, \end{cases} \quad Z_2(i) = \begin{cases} 1 & i = 1, 2, \\ 2 & i = 3, 4, \\ 3 & i = 5, 6. \end{cases}$$

It is easy to see that the matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 6 & 4 & 1 & 5 & 3 \end{pmatrix}^t$$

is an asymmetric U-design that can be represented by the following figure.



Note that in the figure above, each small dashed-line-box contains only one design point.

5.2 Discussion

For a given orthogonal array, there are many corresponding U-designs and apparently these U-designs are not equally preferable as demonstrated in Figures 2 and 4 of Section 3. This raises the question on the choice of U-designs. Keeping in mind the objective of uniformity, two approaches, distance approach and correlation approach, can be utilized. We do not intend to go into details here but merely point out this important problem. The results will be included in a forthcoming paper.

Very often a large number of factors are studied in an experiment. For such an experiment the first problem is to screen out the unimportant factors. Highly fractionated factorial designs are widely used for this purpose(Box, Hunter and Hunter 1978). These designs have the advantages of runsize economy and simplicity in a data analysis. There are also

some disadvantages as described in Section 1. LHDs have been used in screening experiments: Iman and Conover(1980) judges the relative importance of factors by the magnitudes of the linear correlations of the response and the factors by using LHDs; Welch et al.(1992) uses the same designs for screening in their Bayesian approach. As we have mentioned in Section 2, it is possible to use the grouped levels to perform a preliminary data analysis to identify the important factors. However, the main effects defined by the grouped levels are confounded with each other. This raises a very interesting question: how to assess the strengths of confounding?

Acknowledgments

The research was supported by the Natural Sciences and Engineering Research Council of Canada, the Manufacturing Research Corporation of Ontario and the Institute for Improvement in Quality and Productivity. The author is most indebted to Professor Jeff Wu, my supervisor, for his time and energy spent on this work; it is hard to imagine what it would be like without his various supports and valuable comments. Thanks also go to Miss Xiaoqiong Hu for useful discussion and Professor Runchu Zhang for checking the proofs.

REFERENCES

- Bose, R.C. and Bush, K.A.(1952). Orthogonal arrays of strength two and three. *Ann. Math. Statist.* 23, 508–524.

- Box, G.E.P. and Draper, N.R.(1959). A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54, 622–654.
- Box, G.E.P., Hunter, W.G. and Hunter, J.S.(1978). Statistics for experimenters. New York: John Wiley.
- Davis, P.J. and Rabinowitz, P.(1984). Methods of numerical integration, 2th Ed.. Academic Press Inc, Orlando FL.
- Dey, A.(1985). Orthogonal fractional factorial designs. New York: Halsted.
- Iman, R.L. and Conover, W.J.(1980). Small sample sensitivity analysis techniques for computer models, with an application to risk assessment(with discussion). Commun. Statist.–Theory and Methods. A9, 1749–1874.
- Iman, R.L. and Conover, W.J.(1982). A distribution–free approach to inducing rank correlation among input variables. Commun. Statist.–Simula. Computat. 11, 311–334.
- de Launey, W.(1986). A survey of generalized hadamard matrices and difference matrices $D(k, \lambda; G)$ with large k . Utilitas Mathematica. 30, 5–29.
- Mckay, M.D., Conover, W.J. and Beckman, R.J.(1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 21, 239–245.
- Owen, A.(1991). Controlling correlations in latin hypercube samples. J. Amer. Statist. Assoc. To appear.

- Owen, A.(1991). Orthogonal arrays for computer experiments, integration, and visualization. Preprint.
- Plackett, R.L. and Burman, J.P.(1946) The designs of optimum multifactorial experiments. *Biometrika*. 33, 305–325.
- Rao, C.R.(1947). Factorial experiments derivable from combinatorial arrangements of arrays. *Jour. Royal Stat. Soc., Suppl.* 9, 128–139.
- Rao, C.R.(1973). Some combinatorial problems of arrays and applications to design of experiments. In “A Survey of Combinatorial Theory”. Eds. J.N. Srivastava et al., *Amaterdam: North-Holland*, 349–359.
- Rao, J.N.K. and Nigam, A.K.(1991). “Optimal” controlled sampling: a Unified approach. *Biometrika*. To appear.
- Stein, M.(1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*. 29, 143–151.
- Sacks,J. and Ylvisaker,D.(1984). Some model robust designs in regression. *Ann. Statist.* 12, 1324–1348.
- Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., and Morris, M.D.(1992). Screening, predicting and computer experiments. *Technometrics*. To appear.
- Wang, J.C. and Wu, C.F.J.(1991). An approach to the construction of asymmetrical orthogonal arrays. *J. Amer. Statist. Assoc.* 86, 450–456.

Wu, C.F.J.(1989). Construction of $2^m 4^n$ designs via a grouping Scheme. Ann. Statist. 17, 1880–1885.

Wu, C.F.J., Zhang, R. and Wang, R.(1992). Construction of asymmetrical orthogonal arrays. Statistica Sinica. To appear.

Appendix A

We will prove Theorem 1 of Section 4 in this appendix.

Let c_1 be a small cell in \mathcal{C} and L be the large cell that includes c_1 . Suppose that

$$c_1 = A_1 \times \cdots \times A_m \quad \text{and} \quad L = B_1 \times \cdots \times B_m.$$

Let

$$G = \left(\bigcup_{j=1}^m c(A_j) \right) \cup \left(\bigcup_{i \neq j}^m c(B_i \times B_j) \right).$$

By the definition of U-sampling, we first have

$$\text{Prob}(C_2 = c_2 \mid C_1 = c_1) = 0 \text{ for all } c_2 \in G.$$

We proceed to derive the general formula for the expression $\text{Prob}(C_2 = c_2 \mid C_1 = c_1)$ which plays a key role in later development.

Lemma 1

$$\text{Prob}(C_2 = c_2 \mid C_1 = c_1) = \begin{cases} a, & \text{for all } c_2 \subset (\bigcup_{j=1}^m c(B_j))^c = \bigcap_{j=1}^m c(B_j)^c, \\ b, & \text{for all } c_2 \subset (\bigcup_{j=1}^m c(B_j)) \setminus G = \bigcup_{j=1}^m (c(B_j) \setminus G), \end{cases}$$

where a and b are constants. Note that $G^c = (\bigcap_{j=1}^m c(B_j))^c \cup (\bigcup_{j=1}^m (c(B_j) \setminus G))$.

This is obvious from the definition of U-sampling.

Lemma 2

$$\frac{a}{b} = \frac{s - m + 1}{s}.$$

PROOF. Let $c_1 = (u_{11}, u_{12}, \dots, u_{1m})$. Without loss of generality, we may take for

Case (i). $c_2 = (u_{21}, u_{22}, \dots, u_{2m})$ for $c_2 \subset \bigcup_{j=1}^m (c(B_j) \setminus G)$ such that $Z(u_{11}) = Z(u_{21})$ and

$$Z(u_{1j}) \neq Z(u_{2j}), \text{ for } j = 2, \dots, m,$$

Case (ii). $c_2 = (u_{21}^*, u_{22}, \dots, u_{2m})$ for $c_2 \subset \bigcap_{j=1}^m c(B_j)^c$ such that $Z(u_{11}) \neq Z(u_{21}^*)$.

Let $a^* = Z(u_{21}^*)$. To obtain a matrix $U = (u_{ij})$ in Γ_m^2 with the fixed first two rows to be c_1 and c_2 , it is convenient to divide the work into the following three steps.

Step 1. Obtain the elements u_{i1} , $i = 3, \dots, n$, in the first column. The number of ways for doing this is $(n - 2)!$ for both cases.

Step 2. Obtain an orthogonal array $OA(s^2, m, s, 2)$, $A = (a_{ij})$, with the first two rows equal to $Z(c_1)$ and $Z(c_2)$ and the first column fixed by $a_{i1} = Z(u_{i1})$.

Consider the s rows with the elements of the first column to be a^* .

Case (i): For column 2, the number a_{12} must occur once in the s rows above and therefore there are s ways to arrange a_{12} , for column 3, the number a_{13} must occur once in the $s - 1$ rows unoccupied by a_{12} and therefore there are $s - 1$ ways to arrange a_{13} , and so on until a_{1m} is arranged in column m and in the $s - m + 2$ rows unoccupied by a_{12}, \dots, a_{1m-1} . Thus, the total number of ways to arrange a_{12}, \dots, a_{1m} is $s(s - 1) \cdots (s - m + 2)$.

Case (ii): By a similar argument, we can show that the total number of ways to arrange a_{12}, \dots, a_{1m} is $(s - 1)(s - 2) \cdots (s - m + 1)$, since one of the s rows is already occupied by c_2 . After the arrangements above, we claim that the number of ways to obtain the remaining a_{ij} 's is the same for both cases. This is readily seen by first considering the $2s$ rows with the elements of the first column to be a_{11} or a^* , and then considering the remaining $n - 2s$ rows.

Step 3. Obtain u_{ij} , $i \geq 3, j \geq 2$, such that $Z(u_{ij}) = a_{ij}$. It is easily seen that the number of ways of getting such u_{ij} 's is the same for both cases.

In summary, we conclude that the number of matrices in Γ_m^2 with the first two rows equal to c_1 and c_2 is $ks(s - 1) \cdots (s - m + 2)$ for Case (i) and $k(s - 1)(s - 2) \cdots (s - m + 1)$ for Case (ii).

Therefore

$$\frac{a}{b} = \frac{k(s-1)(s-2)\cdots(s-m+1)}{ks(s-1)\cdots(s-m+2)} = \frac{s-m+1}{s}. \quad \square$$

Lemma 3

$$a = \frac{s-m+1}{s^m(s-1)^m(s+1)},$$

$$b = \frac{1}{s^{m-1}(s-1)^m(s+1)}.$$

PROOF. The number of small cells in $c(B_j) \setminus G$ is

$$(n-s)^{m-1}(s-1) = s^{m-1}(s-1)^m \text{ for } j = 1, \dots, m.$$

The number of small cells in $\bigcap_{j=1}^m c(B_j)^c$ is

$$(n-s)^m = s^m(s-1)^m.$$

So we have

$$as^m(s-1)^m + mbs^{m-1}(s-1)^m = 1.$$

By combining this with Lemma 2, we obtain Lemma 3. \square

Let ν be the uniform measure on G^c . Let $g(x | C_1 = c_1)$ be the probability density function of X_2 conditionally on $C_1 = c_1$ with respect to ν . Then we have

Lemma 4

$$g(x | C_1 = c_1) = \begin{cases} \frac{(s+m)(s-m+1)}{s(s+1)}, & \text{for } x \in \bigcap_{j=1}^m c(B_j)^c, \\ \frac{s+m}{s+1}, & \text{for } x \in \bigcup_{j=1}^m (c(B_j) \setminus G). \end{cases}$$

PROOF. It is easy to show that

$$\nu\left(\bigcap_{j=1}^m c(B_j)^c\right) = \frac{s}{s+m},$$

$$\nu\left(\bigcup_{j=1}^m (c(B_j) \setminus G)\right) = \frac{1}{s+m}, \quad j = 1, \dots, m.$$

By Lemma 1, we have

$$g(\mathbf{x} \mid C_1 = c_1) = \begin{cases} \alpha, & \text{for } \mathbf{x} \in \bigcap_{j=1}^m c(B_j)^c, \\ \beta, & \text{for } \mathbf{x} \in \bigcup_{j=1}^m (c(B_j) \setminus G), \end{cases}$$

where α and β are constants. By Lemma 3

$$\text{Prob}(C_2 \subset \bigcap_{j=1}^m c(B_j)^c) = s^m (s-1)^m a = \frac{s-m+1}{s+1}.$$

On the other hand,

$$\begin{aligned} \text{Prob}(C_2 \subset \bigcap_{j=1}^m c(B_j)^c) &= \text{Prob}(X_2 \in \bigcap_{j=1}^m c(B_j)^c) \\ &= \int_{\bigcap_{j=1}^m c(B_j)^c} g(\mathbf{x} \mid C_1 = c_1) d\nu = \alpha \nu(\bigcap_{j=1}^m c(B_j)^c) = \alpha \frac{s}{s+m}. \end{aligned}$$

So

$$\alpha = \frac{(s+m)(s-m+1)}{s(s+1)}.$$

Similarly we have

$$\beta = \frac{s+m}{s+1}.$$

This proves Lemma 4. \square

PROOF OF THEOREM 1: We have

$$\text{Cov}(f(X_1)f(X_2)) = \text{E}(f(X_1)f(X_2)) = \text{E}(\text{E}(f(X_1)f(X_2) \mid C)),$$

where $C = A_1 \times \cdots \times A_m$ and $L = B_1 \times \cdots \times B_m$ are the small and the large cells to which X_1 belongs respectively. Further let

$$G = \left(\bigcup_{j=1}^m c(A_j) \right) \cup \left(\bigcup_{i \neq j}^m c(B_i \times B_j) \right)$$

as before. Then conditionally on C , X_1 and X_2 are independent and X_2 has the probability density function

$$g(\mathbf{x} \mid C) = \begin{cases} \frac{(s+m)(s-m+1)}{s(s+1)}, & \text{for } \mathbf{x} \in \bigcap_{j=1}^m c(B_j)^c, \\ \frac{s+m}{s+1}, & \text{for } \mathbf{x} \in \bigcup_{j=1}^m (c(B_j) \setminus G). \end{cases}$$

Therefore

$$\mathbf{E}(f(X_1)f(X_2)) = \mathbf{E}(\mathbf{E}(f(X_1) | C)\mathbf{E}(f(X_2) | C))$$

and

$$\begin{aligned} & \mathbf{E}(f(X_2) | C) \\ &= \int_{G^c} f(\mathbf{x})g(\mathbf{x} | C) d\nu \\ &= \int_{\bigcup_{j=1}^m (c(B_j) \setminus G)} f(\mathbf{x})g(\mathbf{x} | C) d\nu + \int_{\bigcap_{j=1}^m c(B_j)^c} f(\mathbf{x})g(\mathbf{x} | C) d\nu \\ &= \int_{\bigcup_{j=1}^m (c(B_j) \setminus G)} f(\mathbf{x})[1 + \frac{m-1}{s+1}] d\nu + \int_{\bigcap_{j=1}^m c(B_j)^c} f(\mathbf{x})[1 - \frac{m(m-1)}{s(s+1)}] d\nu \\ &= \int_{G^c} f(\mathbf{x}) d\nu + \frac{m-1}{s+1} \int_{\bigcup_{j=1}^m (c(B_j) \setminus G)} f(\mathbf{x}) d\nu - \frac{m(m-1)}{s(s+1)} \int_{\bigcap_{j=1}^m c(B_j)^c} f(\mathbf{x}) d\nu. \end{aligned}$$

Noting that ν has a density $1+o(1)$ with respect to the Lebesgue measure on G^c , we have

$$\int_{\bigcap_{j=1}^m c(B_j)^c} f(\mathbf{x}) d\nu = \int_{\bigcap_{j=1}^m c(B_j)^c} f(\mathbf{x})[1 + o(1)] d\mathbf{x} = o(1).$$

The last equality follows from the assumption $\mu = 0$. Moreover, it is easy to show that

$$\frac{m-1}{s+1} \int_{\bigcup_{j=1}^m (c(B_j) \setminus G)} f(\mathbf{x}) d\nu = \frac{m-1}{n} \sum_j \mathbf{E}(f(X_1) | c(B_j)) + o(\frac{1}{n}).$$

Also

$$\begin{aligned} & \int_{G^c} f(\mathbf{x}) d\nu = - \int_G f(\mathbf{x}) d\mathbf{x} + o(\frac{1}{n}) \\ &= -\frac{1}{n} \sum_j \mathbf{E}(f(X_1) | c(A_j)) - \frac{1}{n} \sum_{i < j} \mathbf{E}(f(X_1) | c(B_i \times B_j)) + o(\frac{1}{n}). \end{aligned}$$

Therefore we obtain

$$\begin{aligned} \mathbf{E}(f(X_2) | C) &= -\frac{1}{n} \sum_j \mathbf{E}(f(X_1) | c(A_j)) \\ &\quad - \frac{1}{n} \sum_{i < j} \{ \mathbf{E}(f(X_1) | c(B_i \times B_j)) - \mathbf{E}(f(X_1) | c(B_i)) - \mathbf{E}(f(X_1) | c(B_j)) \} + o(\frac{1}{n}). \end{aligned}$$

Note that

$$\begin{aligned} \mathbf{E}[\mathbf{E}(f(X_1) | C)\mathbf{E}(f(X_1) | c(A_j))] &= \text{Var}[\mathbf{E}(f(X_1) | c(A_j))], \\ \mathbf{E}[\mathbf{E}(f(X_1) | C)\mathbf{E}(f(X_1) | c(B_i \times B_j))] &= \text{Var}[\mathbf{E}(f(X_1) | c(B_i \times B_j))], \\ \mathbf{E}[\mathbf{E}(f(X_1) | C)\mathbf{E}(f(X_1) | c(B_j))] &= \text{Var}[\mathbf{E}(f(X_1) | c(B_j))], \end{aligned}$$

and

$$\mathbf{E}[\mathbf{E}(f(X_1) \mid c(B_i))\mathbf{E}(f(X_1) \mid c(B_j))] = 0,$$

$$\mathbf{E}\{\mathbf{E}(f(X_1) \mid c(B_i))[\mathbf{E}(f(X_1) \mid c(B_i \times B_j)) - \mathbf{E}(f(X_1) \mid c(B_i)) - \mathbf{E}(f(X_1) \mid c(B_j))]\} = 0.$$

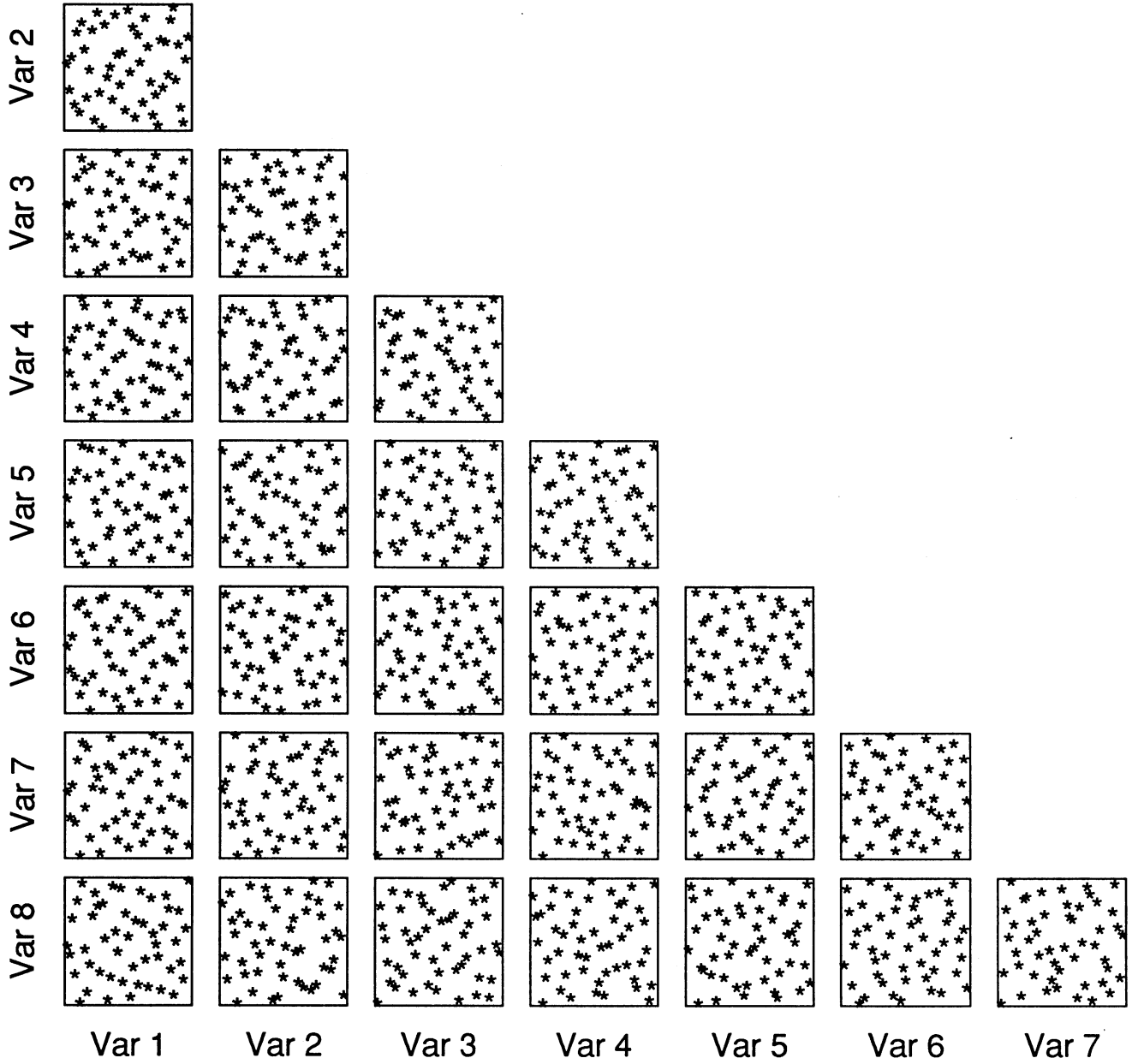
So that

$$\begin{aligned} \text{Cov}(f(X_1), f(X_2)) &= -\frac{1}{n} \sum_j \text{Var}[\mathbf{E}(f(X_1) \mid c(A_j))] \\ &\quad - \frac{1}{n} \sum_{i < j} \text{Var}[\mathbf{E}(f(X_1) \mid c(B_i \times B_j)) - \mathbf{E}(f(X_1) \mid c(B_i)) - \mathbf{E}(f(X_1) \mid c(B_j))] + o\left(\frac{1}{n}\right). \end{aligned}$$

The proof is completed. \square

Appendix B

Random U-design



Random LHD

