J. Ind. Soc. Agril. Statist.
62(1), 2008 : 1-18

Orthogonal Arrays and their Applications[1]

V.K. Gupta
Indian Agricultural Statistics Research Institute, New Delhi

SUMMARY

The purpose of this article is to review method(s) of constructing orthogonal arrays (both symmetric and mixed) by exploiting the concept of resolvable (symmetric) orthogonal arrays and resolvable mixed orthogonal arrays. Several series of symmetric and mixed orthogonal arrays obtained in the literature by using the notion of Kronecker product and Kronecker sum of OAs have also been described. Some general methods of obtaining orthogonal arrays and mixed orthogonal arrays have also been given. Applications of orthogonal arrays, mixed orthogonal arrays and resolvable orthogonal arrays have been described.

Key words: Symmetric and Mixed Orthogonal array; Resolvable orthogonal array; Nested Orthogonal Array; Supersaturated design; Irregular fractions; Balanced repeated replications.

1. INTRODUCTION

Orthogonal Arrays (OAs) were introduced by Rao (1947). An $n \times m$ matrix **A** with entries from a set Σ containing q (≥ 2) symbols is called a (symmetric) *Orthogonal Array* of size n, with m constraints, q symbols, and strength t, if each $n \times t$ sub-matrix of **A** contains all possible q^t row vectors exactly μ times. Such an array is denoted by OA (n, m, q, t). The integer μ is called the index of the OA. Clearly $n = \mu q^t$. When $\mu = 1$, we refer to such arrays as OAs of *index unity*.

Example 1.1 The following is an OA (18, 7, 3, 2), with rows representing the factors (constraints) and columns the runs (size). In this example, n = 18, m = 7, q = 3, t = 2, μ = 2.

0 0 0	1 1 1	2 2 2	0 0 0	1 1 1	2 2 2
0 1 2	0 1 2	0 1 2	0 1 2	0 1 2	0 1 2
0 1 2	1 2 0	2 0 1	2 0 1	0 1 2	1 2 0
0 2 1	1 0 2	2 1 0	1 0 2	2 1 0	0 2 1
0 1 2	1 2 0	1 2 0	0 1 2	2 0 1	2 0 1
0 1 2	2 0 1	0 1 2	1 2 0	2 0 1	1 2 0
0 1 2	0 1 2	2 0 1	1 2 0	1 2 0	2 0 1

For the existence of such arrays, Rao (1947) obtained a bound on the number of runs as given in Theorem 1.1.

Theorem 1.1 The parameters of an OA (n, m, q, t) satisfy the following inequality:

$$n \geq \sum_{j=0}^{u} \binom{m}{j} (q-1)^j, \text{ if } t\,(= 2u, u \geq 1) \text{ is even;}$$

$$n \geq \sum_{j=0}^{u} \binom{m}{j} (q-1)^j + \binom{m-1}{u} (q-1)^{u+1},$$

$$\text{if } t\,(= 2u+1,\ u \geq 0) \text{ is odd.}$$

Rao (1973) defined Mixed Orthogonal Arrays (MOAs) as a generalization of symmetric OAs. A *Mixed Orthogonal Array* (MOA) (n, m, $q_1^{m_1}, q_2^{m_2}, \cdots, q_v^{m_v}$, t) having n rows, m ($\geq$ 2) columns and strength t ($\leq$ m), is an $n \times m$ array, $m = m_1 + m_2 + \ldots + m_v$, in which the first m_1 columns have symbols from {0, 1, 2, ... , $m_1 - 1$}, the next m_2 columns have symbols from {0, 1, 2, ..., $m_2 - 1$}, and so on, and the last m_v columns have symbols from {0, 1, 2, ..., $m_v - 1$}, with the property that every possible *t*-tuple of symbols occurs equally often as rows in every $n \times t$ sub-array.

1 *Keynote Address presented during the 61st Annual Conference of the Indian Society of Agricultural Statistics held at Birsa Agricultural University, Ranchi, during 30 November to 2 December 2007.*

Obviously, $q_i | n \ \forall\ i = 1, 2, ..., v$, where x | y means x divides y. For $q_1 = q_2 = ... = q_v = q$, we get a symmetric OA of strength t.

Example 1.2 The following is an MOA (16, 8, $4^2 \times 2^6$, 2), with rows representing the factors and columns the runs. In this example, n = 16, m = 8, $m_1 = 2$, $m_2 = 6$, $q_1 = 4$, $q_2 = 2$, t = 2.

1	0	1	0	0	1	0	1	0	1	1	0	0	1	1	0
1	0	0	1	1	0	0	1	0	1	0	1	1	1	0	0
1	1	0	0	0	0	1	1	1	1	0	0	0	1	0	1
0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	0
1	0	1	0	1	0	1	0	0	1	1	0	1	0	0	1
0	0	1	1	0	0	1	1	1	1	0	0	1	0	1	0
2	1	0	3	0	3	2	1	0	3	1	2	1	0	2	3
0	3	2	1	3	0	1	2	0	3	1	2	0	1	3	2

Rao bounds on the number of runs in OAs have also been extended to MOAs. To state this bound we define the set $I_f(v)$, where $f \geq 0$ and $v \geq 1$ are integers, by

$$I_f(v) = \{(i_1, i_2, ..., i_v) : i_1 \geq 0, i_2 \geq 0, ..., i_v \geq 0, \sum_{l=1}^{v} i_l = f\}$$

The notation $\sum_{I_f(v)}$ will denote a sum over all v-tuples in $I_f(v)$.

Theorem 1.2 Consider an MOA (n, m, $q_1^{m_1}, q_2^{m_2}, ..., q_v^{m_v}$, t) where without any loss of generality, we assume that $q_1 \leq q_2 \leq ... \leq q_v$. The parameters of the array satisfy

$$n \geq \sum_{v=0}^{u} \sum_{I_f(v)} \binom{m_1}{i_1}\binom{m_2}{i_2} \cdots \binom{m_v}{i_v} (q_1-1)^{i_1}(q_2-1)^{i_2}\cdots(q_v-1)^{i_v},$$
if t (= 2u, u ≥ 1) is even;

$$n \geq \sum_{v=0}^{u} \sum_{I_f(v)} \binom{m_1}{i_1}\binom{m_2}{i_2} \cdots \binom{m_v}{i_v} (q_1-1)^{i_1}(q_2-1)^{i_2}\cdots(q_v-1)^{i_v} + \sum_{I_u(v)} \binom{m_1}{i_1}\binom{m_2}{i_2}\cdots\binom{m_{v-1}}{i_{v-1}}\binom{m_v-1}{i_v} (q_1-1)^{i_1}(q_2-1)^{i_2}\cdots(q_{v-1}-1)^{i_{v-1}}(q_v-1)^{i_v+1},$$
if t (= 2u + 1) is odd, for u ≥ 0.

Henceforth we shall denote a symmetric OA or a MOA as an OA without any distinction. The practical utility of OAs in agricultural, biological, industrial research and quality control work in recent years has given an added impetus to the study of these arrays. A thorough coverage of OAs and related structures is available in the book by Hedayat *et al*. (1999). For the applications of OAs in design of experiments, a reference may be made to Dey and Mukerjee (1999) and Wu and Hamada (2000).

The concept of resolvable and α-resolvable OAs has also been defined (Gupta *et al*. 1982; Dey and Midha, 1996; Hedayat, *et al*. 1999; Sinha *et al*. 2008 and many others). This notion along with the Kronecker Product and Kronecker Sum of OAs generates many series of α-resolvable OAs and consequently MOAs. The concept of nested arrays has also been studied in the literature by many authors (see, *e.g.*, Mukerjee *et al*. (2007) and the references cited therein). These types of arrays are useful for experimentation in two phases; one small experiment of higher accuracy nested in a larger and relatively inexpensive one of lower accuracy.

OAs have many applications. One of the most important applications of OAs is in fractional factorials. Another application has been made in variance estimation of a non-linear statistic from a large scale complex survey data. Special forms of OAs have found applications in computer experiments.

We begin with some preliminaries and give definitions of various concepts in Section 2. In Section 3 we define resolvable OAs. Section 4 describes some methods of construction of OAs and MOAs. Many series of OAs are obtained. Many more can also be obtained from the general methods given. Section 5 describes the applications of OAs in fractional factorials, particularly in generating supersaturated designs. The use of Balanced Repeated Replications (BRR) in variance estimation from a large scale complex survey data has also been shown in this Section. It has been established (see Gupta and Nigam 1987) that BRR are in fact OAs of strength two. A special form of OAs has an interesting application in computer experiments. This has also been mentioned in this Section.

2. PRELIMINARIES

In this Section we give some definitions that would be useful in the sequel. We shall denote a matrix by bold capital letters and vectors by bold small letters. Unless otherwise stated, a vector would be a column vector. We shall denote by $\mathbf{0}_t$ a t-component vector with all elements zero and by $\mathbf{1}_t$ a t-component vector with all elements one. $\mathbf{J}_{mn}$ would denote an m×n matrix with all elements one. Further, $\mathbf{I}_t$ would denote an identity matrix of order t. We shall omit the order of the matrices and the vectors when the orders are obvious.

Most of the techniques for the construction of 2-symbol OAs are special cases of techniques for q-symbol arrays. A structure of 2-symbol OA of interest to us is an *Hadamard Matrix*.

Definition 2.1 An n×n real matrix $\mathbf{H}_n = (h_{ij})$, $h_{ij} = \pm 1$ is an Hadamard matrix of order n *iff* $\mathbf{H}_n\mathbf{H}'_n = \mathbf{H}'_n\mathbf{H}_n = n\mathbf{I}_n$.

Property 1. An Hadamard matrix $\mathbf{H}_n$ of order n exists when n = 1, 2 or n ≡ 0 mod (4), i.e., n = 4w, w being a positive integer. The integer n is also known as an Hadamard number.

Property 2. An Hadamard matrix is invariant with respect to permutation of rows and columns; it is also invariant with respect to multiplication by –1 any row or column.

If the first column of an Hadamard matrix consists of +1's only, then an Hadamard matrix is said to be in *semi-normalized form*. If both first row and first column consist of +1's only, then it is said to be in *normalized form*. Using invariance property, any Hadamard matrix can be written in semi-normalized or normalized form. For online generation of Hadamard matrices of all permissible orders up to 1000, a reference may be made to www.iasri.res.in/design/webhadamard (see also Gupta *et al*. (2007)).

A normalized Hadamard matrix can be written as,

$$\mathbf{H}_n = \begin{bmatrix} 1 & \mathbf{1}'_{n-1} \\ \mathbf{1} & \mathbf{A}_{n-1,n-1} \end{bmatrix}, \quad \mathbf{A}^* = \mathbf{A}_{n,n-1} = \begin{bmatrix} \mathbf{1}' \\ \mathbf{A}_{n-1,n-1} \end{bmatrix}.$$

The (n – 1)×(n – 1) matrix $\mathbf{A}^{@} = \mathbf{A}_{n-1,n-1}$ is called the core of Hadamard matrix $\mathbf{H}_n$.

Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be two matrices of order m × n and p × q respectively.

Definition 2.2 The *Kronecker product* of **A** and **B**, denoted by an mp×nq matrix $\mathbf{F} = \mathbf{A} \otimes \mathbf{B}$, is defined by

$$\mathbf{F} = \left(a_{ij}\mathbf{B}\right)_{1 \le i \le m; 1 \le j \le n}$$

Further, let A and B have entries from a finite additive abelian group G of order q with elements as (0, 1, 2, ..., q – 1).

Definition 2.3 The Kronecker sum of **A** and **B**, denoted by an mp×nq matrix $\mathbf{D} = \mathbf{A} \oplus \mathbf{B}$, is defined by

$$\mathbf{D} = \mathbf{A} \otimes \mathbf{J} + \mathbf{J} \otimes \mathbf{B} = (\mathbf{B}^{a_{ij}})_{1 \le i \le m; 1 \le j \le n},$$

where $\mathbf{B}^{a_{ij}} = (\mathbf{B} + a_{ij}\mathbf{J}) \bmod (q)$.

Let **A** be an OA (n, m, q, t) with symbols as 0, 1, 2, . . . , q – 1. Then we have the following:

Lemma 2.1 For some integer scalar θ, $(\mathbf{A} + \theta\mathbf{J})$ mod (q) is also an OA (n, m, q, t).

Proof. Follows from permutation invariance property of symbols of OAs.

Corollary 2.1.1 $\mathbf{A}_- = (\mathbf{A} - \mathbf{J})$ mod (q) and $\mathbf{A}_+ = (\mathbf{A} + \mathbf{J})$ mod (q) are OAs (n, m, q, t).

For q = 4, the symbols (without loss of generality) are 0, 1, 2, 3. Then symbols of A_- are 0 – 1 = –1 = 3 mod (4); 1 – 1 = 0; 2 – 1 = 1; 3 – 1 = 2. Similarly, the symbols of $\mathbf{A}_+$ are 0 + 1 = 1; 1 + 1 = 2; 2 + 1 = 3; 3 + 1 = 4 = 0 mod (4).

3. RESOLVABLE ORTHOGONAL ARRAYS

In this Section, we define an interesting property of OAs, called resolvable OAs. The use of this notion was first made by Chacko and Dey (1981) though the term resolvable OA was not used by these authors. This was formalized by Gupta *et al*. (1982). Using this concept several families of MOAs of strength two can be obtained.

Definition 3.1 A (symmetric) OA (n, m, q, t), denoted as **A**, is said to be Resolvable Orthogonal Array (ROA) if it is possible to partition A into δ sub-matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_\delta$ such that each $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_\delta$ is an OA (n_1, m, q, t – 1).

Obviously, $n_1 = \upsilon q^{t-1}$. υ is called the index of the array A_i, *i* = 1, 2, . . ., δ. Further, $n_1 = \upsilon q^{t-1} = n/\delta$ and $\delta = \mu q/\upsilon$.

Hedayat *et al*. (1992) defined resolvable OAs of strength two, and exploiting this notion of resolvable OAs obtained several series of MOAs. This has also been given in Dey and Mukerjee (1999) as follows:

Definition 3.2 An OA

$\left(n, m=\sum_{i=1}^{v} m_i, q_1^{m_1}, q_2^{m_2}, \cdots, q_v^{m_v}, 2\right)$ is said to be $(\alpha_1 \times \alpha_2 \times \cdots \times \alpha_v)$-*resolvable* if its rows can be partitioned into $n/\alpha_1 q_1$ sets of $\alpha_1 q_1$ rows each such that for any q_i-symbol column, each possible symbol occurs α_i times within the rows of every set (i = 1, 2, ..., v).

Clearly, for each i = 1, 2, ..., v, $\alpha_i q_i$ does not depend on i and $\alpha_i q_i$ divides n. If $q_1 = q_2 = \ldots = q_v = q$, and $\alpha_1 = \alpha_2 = \cdots = \alpha_v = \alpha$ then an $(\alpha_1 \times \alpha_2 \times \cdots \times \alpha_v)$ -resolvable array reduces to a (symmetric) α-resolvable array.

Example 3.2.1 The following is an $(\alpha_1 \times \alpha_2)$-resolvable MOA (16, 8, $4^2 \times 2^6$, 2), with $\alpha_1 = 1$, $\alpha_2 = 2$ (rows as factors and columns as runs) :

1	0	1	0	0	1	0	1	0	1	1	0	0	1	1	0
1	0	0	1	1	0	0	1	0	1	0	1	1	1	0	0
1	1	0	0	0	0	1	1	1	1	0	0	0	1	0	1
0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	0
1	0	1	0	1	0	1	0	0	1	1	0	1	0	0	1
0	0	1	1	0	0	1	1	1	1	0	0	1	0	1	0
2	1	0	3	0	3	2	1	0	3	1	2	1	0	2	3
0	3	2	1	3	0	1	2	0	3	1	2	0	1	3	2

Dey and Mukerjee (1999) also gave the following theorem :

Theorem 3.1 Let there exist a $\prod_{i=1}^{u} \alpha_i$ – resolvable OA **A**, $\left(n, \sum_{i=1}^{u} m_i, \prod_{i=1}^{u} q_i^{m_i}, 2\right)$ and another OA, **B**, $\left(n/\alpha_1 q_1, \sum_{i=u+1}^{v} m_i, \prod_{i=u+1}^{v} q_i^{m_i}, 2\right)$. Then there exists an OA $\left(n, m=\sum_{i=1}^{v} m_i, \prod_{i=1}^{v} q_i^{m_i}, 2\right)$.

Remark 3.1. A (symmetric) ROA is in fact a nested array in the sense that δ sub-arrays are nested within a big array. We can also denote this as OA ((n, n_1), m, q, (t, t – 1); δ). The bigger array is denoted as **A** and is an OA (n, m, q, t) while the sub arrays are denoted as A_i and each A_i is an OA (n_1, m, q, t – 1) and δ is the number of such sub-arrays. A similar concept can be defined for a $(\alpha_1 \times \alpha_2 \times \ldots \times \alpha_v)$ resolvable array.

Remark 3.2. In this Section we have defined ROA and mentioned in Remark 3.1 that ROAs are a kind of nested arrays. The concept of nested arrays has been studied in the literature in a different sense. We give below the definition of a *Nested Orthogonal Array* (NOA).

Definition 3.3 A Nested Orthogonal Array ((n, n*), m (q, r), t), where r < q and n* < n, is an OA (n, m, q, t) which contains an OA (n*, m, r, t) as a sub-array.

Example 3.3.1 The following is a NOA ((27, 8), 4, (3, 2), 2), with columns forming the n = 27 runs and rows forming the m = 4 factors. In this example, n* = 8, q = 3, r = 2 and t = 2. The complete array **A** is OA (27, 4, 3, 2) and the first 8 columns form the nested sub-array A* (8, 4, 2, 2).

0 0 0 0 1 1 1 1 0 0 0 0 0 1 1 1 1 1 2 2 2 2 2 2 2 2 2

0 0 1 1 0 0 1 1 0 1 2 2 2 0 1 2 2 2 0 0 0 1 1 1 2 2 2

0 1 0 1 0 1 0 1 2 2 0 1 2 2 2 0 1 2 0 1 2 0 1 2 0 1 2

0 1 0 1 1 0 1 0 0 2 2 2 1 2 2 1 2 0 2 2 1 2 1 0 0 0 1

In what follows, the larger array, i.e., the OA (n, m, q, t) would be denoted by A and the smaller array, i.e., the sub-array OA (n*, m, r, t) would be denoted by **A***. Clearly, in a NOA ((n, n*), m(q, r), t), the index, say λ, of the smaller array **A*** cannot exceed the index, say μ, of the larger array **A**. In the example above, λ = 2 and μ = 3. In particular, if λ = μ, then upon removal of **A*** from **A**, one gets an incomplete orthogonal array.

NOAs have practical use in construction of designs when an experiment endeavor consists of two experiments, the expensive one of higher accuracy to be nested in a larger and relatively inexpensive one of lower accuracy. Experimental setups of this kind have been considered among others by Kennedy and O'Hagan (2000), Rees *et al*. (2004), Qian *et al*. (2006), Qian and Wu (2006) and Mukerjee *et al*. (2007).

The following theorem (Mukerjee *et al*. 2007) gives a lower bound to the number of runs of an OA of strength t. This is a generalization of the bounds by Rao (1947) and Bose and Bush (1952).

Theorem 3.2

$$n \ge n^* \sum_{j=0}^{u} \binom{m}{j} \left(r^{-1}q - 1\right)^j, \text{ if } t\left(= 2u, u \ge 1\right) \text{ is even;}$$

$$n \ge n^* \sum_{j=0}^{u} \binom{m}{j} \left(r^{-1}q - 1\right)^j + \binom{m-1}{u} \left(r^{-1}q - 1\right)^{u+1},$$
$$\text{if } t\left(= 2u + 1, u \ge 1\right) \text{ is odd.}$$

The following theorem is also due to Mukerjee *et al*. (2007) :

Theorem 3.3 For the existence of an OA ((n, n*), m, (q, r), t), $t \le k$, it is necessary that

$$k \le \frac{\left(nr^{t-2} - n^* q^{t-2}\right) r}{n^* q^{t-2} \left(q - r\right)} + t - 2.$$

Remark 3.3. We have mentioned in Remark 3.1 that a ROA is in a way NOA. But while comparing a ROA with a NOA it may be noted that in ROA, the bigger array has more strength than the smaller arrays. In NOA, the bigger and smaller arrays have the same strength.

4. CONSTRUCTION OF ORTHOGONAL ARRAYS

The notion of resolvable OAs has been exploited intensely in the literature to develop methods of construction of OAs. The methods developed also use Kronecker product and Kronecker sum of matrices. We give below a review of some methods of construction of OAs of strength two, although some OAs of strength t have also been reported. Some general methods of construction of OAs have also been given.

We shall, throughout, denote by (0, 1, 2, . . ., q – 1) the q symbols of an OA (n, m, q, t) and by (0, 1) the two symbols of an Hadamard matrix unless required to be represented as (+1, –1).

Theorem 4.1 The existence of an Hadamard matrix of order 4w implies the existence of an OA (4w, 4w–1, 2, 2).

Theorem 4.2 If C is an OA ($w2^t$, m, 2, t), where t is even, then the array **B** defined by

$$\mathbf{B} = \begin{bmatrix} \mathbf{C} & \bar{\mathbf{C}} \\ \mathbf{0}' & \mathbf{1}' \end{bmatrix}$$

is an OA ($w2^{t+1}$, m+1, 2, t+1), where $\bar{\mathbf{C}}$ is the array obtained from C by transforming $0 \leftrightarrow 1$.

Corollary 4.2.1 The existence of an Hadamard matrix of order 4w is equivalent to the existence of an OA (8w, 4w, 2, 3).

Let $\mathbf{A}^{@}$ [(n–1)×(n–1)] be the core of an Hadamard matrix of order n = 4w. Then we have the following theorem:

Theorem 4.3 The existence of an Hadamard matrix of order 4w, with symbols as (+1, –1), implies that of symmetric Balanced Incomplete Block (BIB) designs with parameters

(i) v = b = 4w–1, r = k = 2w–1, λ = w–1

(ii) v = b = 4w–1, r = k = 2w, λ = w

Proof. Follows by noting that $\mathbf{N}_1 = \frac{1}{2}\left(\mathbf{11}' + \mathbf{A}^{@}\right)$ and $\mathbf{N}_2 = \frac{1}{2}\left(\mathbf{11}' - \mathbf{A}^{@}\right)$ are incidence matrices of BIB designs with parameters in Theorem 4.3.

Theorem 4.4 Let $\mathbf{N_1}$ and $\mathbf{N_2}$ be as defined in Theorem 4.3. Further, define

$$\mathbf{B} = \begin{bmatrix} \mathbf{1} & \mathbf{N}_1 & \mathbf{N}_2 & \mathbf{0} \\ 0 & \mathbf{0}' & \mathbf{1}' & 1 \end{bmatrix}$$

Then

(i) **[1 $\mathbf{N_1}$]** and **[$\mathbf{N_2}$ 0]** are OA (4w, 4w–1, 2, 2).

(ii) **B** is an OA (8w, 4w, 2, 3).

Example 4.4.1 Let $\mathbf{H}_8 = \begin{bmatrix} 1 & \mathbf{1}' \\ \mathbf{1} & \mathbf{A}^{@} \end{bmatrix}$,

$$\mathbf{A}^{@} = \begin{bmatrix} 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 \end{bmatrix}$$

Then

$$\mathbf{N}_1 = \frac{1}{2}(\mathbf{11}' + \mathbf{A}^{@}) = \begin{bmatrix} 1&1&0&1&0&0&0\\ 0&1&1&0&1&0&0\\ 0&0&1&1&0&1&0\\ 0&0&0&1&1&0&1\\ 1&0&0&0&1&1&0\\ 0&1&0&0&0&1&1\\ 1&0&1&0&0&0&1 \end{bmatrix},$$

$$\mathbf{N}_2 = \frac{1}{2}(\mathbf{11}' - \mathbf{A}^{@}) = \begin{bmatrix} 0&0&1&0&1&1&1\\ 1&0&0&1&0&1&1\\ 1&1&0&0&1&0&1\\ 1&1&1&0&0&1&0\\ 0&1&1&1&0&0&1\\ 1&0&1&1&1&0&0\\ 0&1&0&1&1&1&0 \end{bmatrix}$$

are the incidence matrices of BIB designs with parameters v = b = 7, r = k = 3, λ = 1, and v = b = 7, r = k = 4, λ = 2. Further, [$\mathbf{1}\ \mathbf{N}_1$] and [$\mathbf{N}_2\ \mathbf{0}$] are OA (8, 7, 2, 2). Also, **B**, defined below, is OA (16, 8, 2, 3):

$$\mathbf{B} = \left[\begin{array}{cccccccc|cccccccc} 1&1&1&0&1&0&0&0&0&0&1&0&1&1&1&0\\ 1&0&1&1&0&1&0&0&1&0&0&1&0&1&1&0\\ 1&0&0&1&1&0&1&0&1&1&0&0&1&0&1&0\\ 1&0&0&0&1&1&0&1&1&1&1&0&0&1&0&0\\ 1&1&0&0&0&1&1&0&0&1&1&1&0&0&1&0\\ 1&0&1&0&0&0&1&1&1&0&1&1&1&0&0&0\\ 1&1&0&1&0&0&0&1&0&1&0&1&1&1&0&0\\ \hline 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1 \end{array}\right]$$

We now give below a theorem which can be used to construct a large number of OAs (Gupta *et al*. 1982; Dey and Agrawal, 1985; Dey 1985, pp. 55-57, Dey and Mukerjee 1999, pp. 61-68).

Theorem 4.5 The existence of a (symmetric) ROA, **A** (n, m, q, 2) implies the existence of an OA (n, m + 1, $\delta \times q^m$, 2).

Proof. Let A be a ROA (n, m, q, 2). The array A can be partitioned as

$$\mathbf{A} = \left[\mathbf{A}_1' \ \vdots\ \mathbf{A}_2' \ \vdots\ \mathbf{A}_3' \ \vdots\ \cdots \ \vdots\ \mathbf{A}_\delta'\right]',$$

where n = $q\delta\alpha$. Then the matrix

$$\mathbf{B} = \left[\begin{matrix} \mathbf{A}_1' & \vdots & \mathbf{A}_2' & \vdots & \mathbf{A}_3' & \vdots & \cdots & \vdots & \mathbf{A}_\delta' \\ \mathbf{0}' & \vdots & \mathbf{1}' & \vdots & 2\mathbf{1}' & \vdots & \cdots & \vdots & (\delta-1)\mathbf{1}' \end{matrix}\right]'$$

is the OA($q\delta\alpha$, m + 1, $\delta \times q^m$, 2).

Corollary 4.5.1 (Dey 1985, pp. 55-57, Dey and Mukerjee, 1999, pp. 61-68). The existence of a $\prod_{i=1}^{u}\alpha_i$ -resolvable OA $\left(n, m = \sum_{i=1}^{v} m_i, \prod_{i=1}^{v} q_i^{m_i},\ 2\right)$ implies that of a $\left(n, m = \sum_{i=1}^{v} m_i + 1,\ (n/\alpha_1 q_1) \times \prod_{i=1}^{v} q_i^{m_i},\ 2\right)$.

Example 4.5.1.1 From the $(\alpha_1 \times \alpha_2)$ resolvable OA (16, 8, $4^2 \times 2^6$, 2), with $\alpha_1 = 1$, $\alpha_2 = 2$ in Example 3.3.1, we can obtain an OA (16, 9, $4^3 \times 2^6$, 2) as follows (rows as factors and columns as runs):

1	0	1	0	0	1	0	1	0	1	1	0	0	1	1	0
1	0	0	1	1	0	0	1	0	1	0	1	1	1	0	0
1	1	0	0	0	0	1	1	1	1	0	0	0	1	0	1
0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	0
1	0	1	0	1	0	1	0	0	1	1	0	1	0	0	1
0	0	1	1	0	0	1	1	1	1	0	0	1	0	1	0
2	1	0	3	0	3	2	1	0	3	1	2	1	0	2	3
0	3	2	1	3	0	1	2	0	3	1	2	0	1	3	2
0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3

Application of Theorem 4.5 has been made in the literature to obtain several families of OAs.

Series 4.5.1 ~ OA (λq^2, $\lambda q + 1$, $\lambda q \times q^{\lambda q}$, 2)

Bose and Bush (1952) constructed a series of OAs (λq^2, λq, q, 2), where λ and q are powers of the same prime. This OA is resolvable with $\delta = \lambda q$. Using Theorem 4.5 gives the required OA.

From this series of OA, we can also obtain, using the principle of replacement, a series of OA (λq^2, $\lambda q + 2$, $\lambda \times q \times q^{\lambda q}$, 2).

Series 4.5.2 ~ OA ($2q^t$, m, $2q \times q^{m-1}$, 2)

Addelman and Kempthorne (1961) constructed OA ($2q^t$, m, q, 2), where q is a prime power, t ($\geq$ 2) is a positive integer, and m = $\frac{2(q^t - 1)}{q-1} - 1$. This is a

q^{t-2}-resolvable OA for some (m – 1) factors with $\delta = 2q$, and each set containing q^{t-1} runs (Gupta *et al.* 1982). Using Theorem 4.5 gives the required OA.

From this series of OA, we can also obtain, using the principle of replacement, a series of MOA ($2q^t$, m + 1, $2 \times q^m$, 2).

Series 4.5.3 ~ OA (nq, q(n – 1) + 1, $q \times 2^{q(n-1)}$, 2)

Consider an Hadamard matrix $\mathbf{H_n}=[\mathbf{1}\ \mathbf{A}^*]$ of order $n \geq 4$. Select another Hadamard matrix $\mathbf{H}_q$ of order q. Obtain $\mathbf{A} = \mathbf{H}_q \otimes \mathbf{A}^*$. The matrix **A** is

$$\mathbf{A} = \begin{bmatrix} \pm\mathbf{A}^* & \cdots & \pm\mathbf{A}^* & \cdots & \pm\mathbf{A}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \pm\mathbf{A}^* & \cdots & \pm\mathbf{A}^* & \cdots & \pm\mathbf{A}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \pm\mathbf{A}^* & \cdots & \pm\mathbf{A}^* & \cdots & \pm\mathbf{A}^* \end{bmatrix}$$

A is a resolvable OA (nq, q(n – 1), 2, 2) with $\delta = q$. Using Theorem 4.5 gives the required OA.

The following theorem follows from Theorem 1 in Sinha *et al*. (2008). This theorem together with Theorem 4.5 helps in generating a large number of OAs.

Theorem 4.6 If $\mathbf{A} = (a_{ij})$ is an OA (n, m, q, t) and $\mathbf{B} = (b_{ij})$ is another OA (n_1, m_1, q_1, t_1), then

$$\mathbf{A} \oplus \mathbf{B} = \mathbf{D} = (\mathbf{B} + a_{ij}\mathbf{J}) \ \bmod\ (q^*)$$

is an OA (nn_1, mm_1, q_1, p), where $t \leq t_1$, and p = 2 if t, t_1 = 2 and p = 3 if max(t, t_1) =3.

Corollary 4.6.1 If **A** is an Hadamard matrix of order n = 4w and **B** as in Theorem 4.6, then $\mathbf{A} \oplus \mathbf{B}$ is an OA (nn_1, nm_1, q_1, 2).

Corollary 4.6.2 Let $\mathbf{H_n} = [\mathbf{1}\ \mathbf{A}^*]$ be an Hadamard matrix of order n = 4w and **B** as in Theorem 4.6. Then $\mathbf{A}^* \oplus \mathbf{B}$ is a resolvable OA (nn_1, (n–1)m_1, q_1, 2). From this one can also obtain, using Theorem 4.5, an OA $\left(nn_1,\ (n-1)m_1 + 1,\ n \times q_1^{(n-1)m_1},\ 2\right)$.

Using Theorem 4.6 and Corollary 4.6.1, we can obtain many series of OAs.

We now give a generalization of Theorem 4.6. Let **A** be an OA (n, m, q, t). Let **B** be a MOA $\left(n^*,\ m^*,\ s_1^{m_1} \times s_2^{m_2} \times \cdots \times s_g^{m_g},\ t^*\right)$;

$m_1 + m_2 + ... + m_g = m^*$, $t < t^*$.

We may write $\mathbf{B} = [\mathbf{B}_1\ \ \mathbf{B}_2\ \ ...\ \mathbf{B}_l\ ...\ \mathbf{B}_g]$. We then give the following definition:

Definition 4.1 Let **A** and **B** be as defined above. Then

$$\mathbf{A} \oplus \mathbf{B} = \mathbf{D} = \Big[\left(\mathbf{B}_1 + a_{ij}\mathbf{J}\right)_{1\leq i\leq n^*; 1\leq j\leq m_1} \bmod(s_1) \ \ldots \left(\mathbf{B}_\ell + a_{ij}\mathbf{J}\right)_{1\leq i\leq n^*; 1\leq j\leq m_\ell} \bmod(s_\ell) \ \ldots \left(\mathbf{B}_g + a_{ij}\mathbf{J}\right)_{1\leq i\leq n^*; 1\leq j\leq m_g} \bmod(s_g) \Big]$$

We then have the following theorem :

Theorem 4.7 Let **A** be an OA (n, m, q, t) and **B** be a MOA $\left(n^*,\ m^*,\ s_1^{m_1} \times s_2^{n_2} \times \cdots \times s_g^{m_g},\ t^*\right)$;

$m_1 + m_2 + \cdots + m_g = m^*$. Then $\mathbf{A} \oplus \mathbf{B}$ is a

MOA $\left(nn^*,\ mm^*,\ s_1^{mm_1} \times s_2^{mm_2} \times \cdots \times s_g^{mm_g},\ p\right)$, $t \leq t^*$, and p = 2 if t, t^* = 2 and p = 3 if max(t, t*) = 3.

Using two-symbol OA **A** (n, n–1, 2, 2) in Theorem 4.6 and making different choices of OA **B** gives many series of MOAs.

5. APPLICATIONS OF OAs

OAs have many applications in different fields. One of the most important applications of OAs is in fractional factorial plans. We begin with some definitions.

Definition 5.1 A fractional factorial plan is said to be of *resolution* 2t + 1 if it permits orthogonal estimation of mean and all effects up to t factors assuming that all effects of t + 1 factors or more are absent. Similarly, a fractional factorial plan is said to be of *resolution* 2t if it permits the orthogonal estimation of mean and all effects up to t – 1 factors assuming that all effects of t + 1 factors or more are absent.

An *Orthogonal Main Effect Plan* (OMEP) is a fractional factorial plan that permits orthogonal estimation of mean and main effects assuming that all two-factor and higher order interactions are absent. In an OMEP, the best linear unbiased estimates of contrasts belonging to different main effects are uncorrelated. It then follows that OAs of strength two are OMEPs but the converse is not true.

Further, an OA of strength t implies the existence of a resolution t + 1 plan. The converse, however, may not be true. Similarly, an MOA of strength two implies the existence of an OMEP for asymmetrical factorial experiment. The converse, however, may not be true.

We then have the following theorem:

Result 5.1 All the OAs and MOAs of strength t constructed in Section 4 are Resolution t+1 plans.

Result 5.2 All the OAs and MOAs of strength two constructed in Section 4 are OMEPs.

In an MOA of strength two every pair of symbols corresponding to any two columns appears a constant number of times. In the context of asymmetric factorials, this combination balance provides a measure of factor orthogonality. It may be noted that combination balance implies factor orthogonality, though the converse is not true.

It is also true that an MOA of strength t is also an MOA of strength t – 1. In that sense in an MOA of strength two, every column would have every symbol appearing a constant number of times. In the context of asymmetric factorials, it means that in the design with n runs all the q_j levels (or symbols) of the j^{th} factor (or column), j = 1, 2, ..., m, appear a constant number of times. We shall call such a design as *level-balanced* or *simply balanced*. In the context of Supersaturated Designs (SSDs), the concept of balance has been used for two-level factor designs. In the case of multi-level and mixed-level SSDs, this concept has been defined as a U-type design. But we shall maintain a uniform definition and call all such designs as balanced. We shall denote by $B(n; q_1, q_2, ..., q_m)$ a class of balanced designs with specified parameters.

5.1 Supersaturated Design

In agricultural, biological and industrial experiments, there occur experimental situations where a large number of factors are to be tested but only few of the factors are active. In such experiments, the experimenter's endeavor is to minimize the number of experiments (or runs or treatment combinations) to identify the active factors for efficient utilization of resources and minimization of cost and time. A *Supersaturated Design* (SSD) is essentially a fractional factorial design in which the degrees of freedom for all its main effects and the intercept term exceed the total number of design runs. SSDs are also fractional factorial designs of which the numbers of columns for allocating factors is greater than those for ordinary orthogonal designs. Because of their run size economy, these designs can be broadly exploited to screen active factor main effects when experimentation is expensive and the number of factors is large. The usefulness of such a supersaturated design relies upon the realism of *effect sparsity*, namely, that the number of dominant active factors is small and the total number of factors in the experiment is very large. Further, one does not know how many, or which of the factors are active. The goal is to identify these active factors with so-called screening experimentation.

Suppose that an experimenter is interested in a fractional factorial design in m factors with the j^{th} factor having q_j levels, j = 1, 2, ..., m. Suppose further that the experimenter can afford only $n \le \Pi_{j=1}^{m} q_j$ experiments (or treatment combinations or runs). Define $\upsilon = \frac{\sum_{j=1}^{m}(q_j - 1)}{n-1}$. A fractional factorial design is called *saturated* if $\upsilon = 1$. The design is called *supersaturated* if $\upsilon > 1$.

A distinctive aspect of SSDs is the measure of non-orthogonality between any two columns of the design. Typically, the criteria of orthogonality (or non-orthogonality) in the literature measure closeness (or departure) to being an OA / MOA of strength two. We describe briefly various measures of the non-orthogonality between two columns in the sequel.

Let $\mathbf{X} = ((X_{ij}))$ be an $n \times m$ design matrix of a factorial experiment with m factors and n runs. For a two-level design, the levels are generally denoted by +1 and –1, and so in that case X_{ij} = +1 or –1. The design matrix $\mathbf{X}$ is called column-orthogonal if $\mathbf{X'X}$ is a diagonal matrix. Let s_{ij} be the (i, j)th element of $\mathbf{X'X}$. The popular $E(s^2)$ criterion, proposed by Booth and Cox (1962), is to minimize

$$E\left(s^2\right) = \sum_{1\le i<j\le m} s_{ij}^2 \Big/ \binom{m}{2}. \tag{5.1}$$

The lower bound to $E(s^2)$, obtained by Nguyen (1996), is

$$E\left(s^2\right) \ge \frac{n^2(m-n+1)}{(m-1)(n-1)} = L[E(s^2)]. \tag{5.2}$$

The same bound was obtained independently by Tang and Wu (1997). Better bounds are now available (see *e.g*., Butler *et al*. 2001; Bulutoglu and Cheng 2004; Ryan and Bulutoglu 2007; and Das *et al.* 2007).

$E(s^2)$ criterion is meaningful when we are trying to measure the closeness of SSD to an OA of strength two in two symbols. However, for multi-level or mixed-level SSDs, to measure the closeness of the design to OA / MOA of strength two, we measure factor orthogonality via combination balance.

For a multi-level $(n; q^m)$ – SSD **X**, let $x^1, x^2, \ldots, x^m$ be its m columns, q(>2). For every pair of columns(x^i, x^j), Lu and Sun (2001) defined a measure of departure from orthogonality for two columns $\mathbf{x}^i$ and $\mathbf{x}^j$ as

$$d_{ij}^2 = \sum_{u=1}^{q}\sum_{v=1}^{q}\left[n_{uv}^{(ij)} - \frac{n}{q^2}\right]^2 \tag{5.3}$$

where $n_{uv}^{(ij)}$ is the number of (u, v)-pairs in $(\mathbf{x}^i, \mathbf{x}^j)$, and n/q^2 stands for the average frequency of the level-combinations in each pair of columns $\mathbf{x}^i$ and $\mathbf{x}^j$.

Based on these d_{ij}^2, two criteria are defined as global measures of departure from orthogonality of the design.

The first criterion is the average of d_{ij}^2's for all the column-pairs in a SSD, i.e.

$$E\left(d^2\right) = \sum_{i<j} d_{ij}^2 \Big/ \binom{m}{2} \tag{5.4}$$

The second criterion is $\max\left(d^2\right) = \max_{i<j} d_{ij}^2$.

For a mixed level - SSD **X**, let $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m$ be its m columns. Fang *et al*. (2000) generalized d_{ij}^2 to

$$f(\mathbf{x}^i, \mathbf{x}^j) = \sum_{u=1}^{q_i}\sum_{v=1}^{q_j}\left|n_{uv}^{(ij)} - \frac{n}{q_i q_j}\right|, \tag{5.5}$$

q_i and q_j are the number of levels in the columns $\mathbf{x}^i$ and $\mathbf{x}^j$, respectively and n/q_iq_j stands for the average frequency of level-combination in each pair of columns x^i and x^j. They also introduced the concept of U-type design. **X** is called a U-type design in the class $U(n; q_1, q_2, \ldots, q_m)$ if it is level balanced. Further a U-type design is called an orthogonal design, if it is an MOA of strength two.

Fang *et al*.(2000) also defined the following criteria for measuring combination non-orthogonality of the design **X**

1. $\mathrm{Ave}|f| = \sum_{1\le i<j\le m} f\left(\mathbf{x}^i, \mathbf{x}^j\right) \Big/ \binom{m}{2}$
2. $\mathrm{Ave}\left(f^2\right) = \sum_{1\le i<j\le m} f\left(\mathbf{x}^i, \mathbf{x}^j\right)^2 \Big/ \binom{m}{2}$ (5.6)
3. $f_{max} = \max_{1\le i<j\le m} f(\mathbf{x}^i, \mathbf{x}^j)$
4. N f_{max} = The frequency of $\{f(\mathbf{x}^i, \mathbf{x}^j) = f_{max}\}$
5. N_{non-od} = The number of $\{f(\mathbf{x}^i, \mathbf{x}^j) \neq 0\}$

The criteria Ave|f| and $\mathrm{Ave}(f^2)$ give the combination imbalance of all column pairs of **X** in the average sense. The criterion f_{max} shows the worst imbalance in all pairs of columns of **X**. Fang *et al*. (2000) also showed that for a two-level design, $f(\mathbf{x}^i, \mathbf{x}^j) = |s_{uv}|$, Ave | f | = Ave | s | and $\mathrm{Ave}(f^2) = E(s^2)$.

Fang *et al*. (2003) used the idea of $\mathrm{Ave}(f^2)$ to define a measure of column non-orthogonality for asymmetrical SSDs and called it as $E(f_{NOD})$. Fang *et al*. (2003) also obtained a lower bound to $E(f_{NOD})$. The results are given below:

$$f_{NOD}^{ij} = \sum_{u=1}^{q_i}\sum_{v=1}^{q_j}\left[n_{uv}^{(ij)} - \frac{n}{q_i q_j}\right]^2 \tag{5.7}$$

Here the subscript NOD stands for non-orthogonality of the design.

A criterion $E(f_{NOD})$ is defined as minimizing

$$E\left(f_{NOD}\right) = \sum_{1 \le i < j \le m} f_{NOD}^{ij} \Big/ \binom{m}{2} \quad (5.8)$$

For an orthogonal array, $E(f_{NOD}) = 0$. Theorem 1 of Fang *et al*. (2003) gave a lower bound to $E(f_{NOD})$. A sharper bound than this was obtained by Fang *et al*. (2004) and is given below in Theorem 5.1.1.

Theorem 5.1.1 For any design $\mathbf{X} \in B(n; q_1, q_2, \cdots, q_m)$ we have

$$E\left(f_{NOD}\right) \ge \frac{n(n-1)}{m(m-1)}\left[(\gamma+1-\psi)(\psi-\gamma)+\psi^2\right]$$

$$+ C\left(n, q_1, q_2, \cdots, q_m\right) = L\left[E\left(f_{NOD}\right)\right] \quad (5.9)$$

where $C\left(n; q_1, q_2, \cdots, q_m\right) = \frac{nm}{m-1} - \frac{1}{m(m-1)}$

$$\left(\sum_{i=1}^{m} \frac{n^2}{q_i} + \sum_{i,j=1, j \ne i}^{m} \frac{n^2}{q_i q_j}\right)$$

depends on **X** only through n, q_1, q_2, ..., q_m.

Here for any design $\mathbf{X} \in B(n; q_1, q_2, \cdots, q_m)$,

$\psi = \frac{\sum_{i=1}^{m} n/q_i - m}{(n-1)}$, $\gamma = [\psi]$, and $[x]$ denotes the integer part of x.

For a mixed level $(n; q_1, q_2, ..., q_m)$ – SSD X, let $\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^m$ be its m columns. For every pair of columns $(\mathbf{x}^i, \mathbf{x}^j)$, Yamada and Lin (1999) defined the following χ^2-value:

$$\chi^2\left(\mathbf{x}^i, \mathbf{x}^j\right) = \frac{q_i q_j}{n} \sum_{u=1}^{q_i} \sum_{v=1}^{q_j} \left(n_{uv}^{ij} - \frac{n}{q_i q_j}\right)^2 \quad (5.10)$$

Obviously, the value of $\chi^2(\mathbf{x}^i, \mathbf{x}^j)$ just measures the non-orthogonality between two columns $\mathbf{x}^i$ and $\mathbf{x}^j$. Then the $E\left(\chi^2\right)$ value defined as

$$E\left(\chi^2\right) = \sum_{1 \le i < j \le m} \chi^2\left(\mathbf{x}^i, \mathbf{x}^j\right) / \binom{m}{2} \quad (5.11)$$

can be used to evaluate the overall non-orthogonality between the columns of **X**. A SSD is called $E(\chi^2)$ optimal if it minimizes the value of $E(\chi^2)$. Ai, Fang and He (2007) have obtained a lower bound to the value of $E(\chi^2)$ for a balanced design.

Theorem 5.2.2 For any design $X \in B(n; q_1, q_2, \cdots, q_m)$,

$$E\left(\chi^2\right) \ge \frac{1}{m(m-1)(n-1)}\left(nm - \sum_{k=1}^{m} q_k\right)^2 +$$

$$C_1\left(n; q_1, q_2, \cdots, q_m\right) = L\left[E\left(\chi^2\right)\right] \quad (5.12)$$

where

$$C(n; q_1, q_2, ..., q_m) = \frac{1}{m(m-1)}\left[\left(\sum_{k=1}^{m} q_k\right)^2 - n\sum_{k=1}^{m} q_k\right] - n$$

Ai *et al.* (2007) further state that any saturated mixed orthogonal array of strength two is $E(\chi^2)$ – optimal.

5.1.1 Supersaturated Designs from Resolvable Orthogonal Arrays

Gupta *et al*. (2007) gave a method of obtaining SSDs for asymmetrical factorial experiments that have intuitive appeal in the sense that these are having high efficiency in terms of criteria defined above. The method of construction of SSDs is as follows:

Method 5.1.1 Consider a ROA (OA), A, (n, m, q, 2) of index unity. The array **A** can be partitioned as

$A = \left[\mathbf{A}_1' \vdots \mathbf{A}_2' \vdots \mathbf{A}_3' \vdots \cdots\cdots \vdots \mathbf{A}_\delta'\right]'$, where $\delta = \frac{n}{q}$. Obviously, $\mathbf{A}_i$ matrix is such that every column of $\mathbf{A}_i$ has all the q-symbols appearing once i = 1, 2, ..., δ.

For some positive integer t < δ, select t sub-matrices of **A**. Let the chosen sub-matrices be denoted as $\mathbf{A}_{(1)}$, $\mathbf{A}_{(2)}$, . . ., $\mathbf{A}_{(t)}$ Consider the matrix **B** given by

$$\mathbf{B} = \begin{bmatrix} \mathbf{0}' & \mathbf{1}' & 2\mathbf{1}' & \cdots & (t-1)\mathbf{1}' \\ \mathbf{A}_{(1)}' & \mathbf{A}_{(2)}' & \mathbf{A}_{(3)}' & \cdots & \mathbf{A}_{(t)}' \end{bmatrix}'$$

The rows of **B** then form the runs of a SSD for $t.q^m//qt$ experiment. If t = q, then we get a supersaturated design for $q^{m+1}//q^2$. Indeed, it is possible to generate $\binom{\delta}{t}$ such designs. The properties of these $\binom{\delta}{t}$ designs are discussed in Section 5.

Using different choices of ROA A, Gupta *et al*. (2007) obtained several series of mixed levels SSDs. The series of SSDs are (i) $t.q^{\lambda q}//qt$, $t < \lambda q$, λ and q are powers of the same prime, (ii) $t.q^{m-1}//t.q$, $2 \le t \le \delta$, q is a prime power, (iii) $t.2^{p(n-1)}//tn$, p, $t \le w$, w and n are Hadamard numbers.

5.1.2 Design Evaluation

Using Method 5.1.1 on various series of mixed level SSDs Gupta *et al*. (2007) generated a large number of SSDs for asymmetrical factorials. Theorems 1 and 2 give the lower bounds $L[E(f_{NOD})]$ and $L[E(\chi^2)]$ to $E(f_{NOD})$ and $E(\chi^2)$. For a design d generated we obtain the values of $E(f_{NOD})$ and $E(\chi^2)$ as respectively, $E_d(f_{NOD})$ and $E_d(\chi^2)$. We then define f_{NOD}-efficiency and χ^2-efficiency as

$$f_{NOD}\text{-efficiency} = \frac{L\left[E\left(f_{NOD}\right)\right]}{E_d\left(f_{NOD}\right)},$$

$$\chi^2\text{-efficiency} = \frac{L\left[E\left(\chi^2\right)\right]}{E_d\left(\chi^2\right)} \quad (5.13)$$

A design is $f_{NOD}(\chi^2)$-optimal when $f_{NOD}(\chi^2)$-efficiency is equal to 1.00. A design with high efficiency is acceptable design. All the designs generated along with their $f_{NOD}(\chi^2)$-efficiency are available at www.iasri.res.in/design/ in the link 'Supersaturated Design / Catalogue.'

5.2 Fractional Factorial

Consider an experimental situation in which all main effects of factors are to be estimated through a fractional factorial design. Besides, some, and not all, two-factor interactions present also need to be estimated through the design.

Three quarter replicates of 2^m designs, $m \ge 3$, ¾(2^m), series, is a special class of fractional factorial designs. In regular fractional 2^m designs, number of runs is an integral power of two, i.e., 4, 8, 16, 32, 64, ... The ¾(2^m) series, obtained by combination of three sets of fractions, provide irregular fractional designs in 6, 12, 24, 48, 96, … runs.

Three-quarter (¾) designs are two-level fractional factorial designs that require only three-quarters of the number of runs of the original design. Three-quarter fractional factorial designs help save resources in two different contexts. Firstly, an experimenter may wish to perform additional runs after having completed a fractional factorial, so as to de-alias certain specific interaction patterns. Secondly, the experimenter may wish to use a (¾) design to begin with and save (¼) of the run requirement of a regular design.

Suppose that an experiment involves investigating four experimental factors, viz., X_1, X_2, X_3 and X_4, and we have designed and run a 2^{4-1} fractional factorial design. The 2^{4-1} design is of resolution IV, which allows orthogonal estimation of mean and all main effects in the presence of two factor interactions, i.e., the main effects are not aliases of any main effect and are confounded with, at worst, three-factor interactions, and two-factor interactions are confounded with other two factor interactions. The design matrix, in standard order, is shown in Table 5.2.1 along with all the two-factor interaction columns. Note that the column for X_4 is constructed by multiplying columns for X_1, X_2, and X_3 together (i.e., $X_4 = X_1X_2X_3$).

Table 5.2.1. The 2^{4-1} design plus 2-factor interaction columns shown in standard order (Note that $X_4 = X_1X_2X_3$)

Run	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3	X_2X_4	X_3X_4
1	–1	–1	–1	–1	1	1	1	1	1	1
2	1	–1	–1	1	–1	–1	1	1	–1	–1
3	–1	1	–1	1	–1	1	–1	–1	1	–1
4	1	1	–1	–1	1	–1	–1	–1	–1	1
5	–1	–1	1	1	1	–1	–1	–1	–1	1
6	1	–1	1	–1	–1	1	–1	–1	1	–1
7	–1	1	1	–1	–1	–1	1	1	–1	–1
8	1	1	1	1	1	1	1	1	1	1

Note also that $X_1X_2 = X_3X_4$, $X_1X_3 = X_2X_4$, and $X_1X_4 = X_2X_3$. These follow from the generating relationship $X_4 = X_1X_2X_3$ and tells us that we cannot estimate any two-factor interaction that is free of some other two-factor alias.

Suppose that the experimenter is interested in estimating all two-factor interactions that involved factor X_1; that is, we want to estimate interactions X_1X_2, X_1X_3, and X_1X_4 free of two-factor confounding.

One way of doing this is to run the `other half' of the design, i.e., additional eight runs formed from the relationship $X_4 = -X_1X_2X_3$. Putting these two 'halves' together—the original one and the new one, gives a 2^4 design in sixteen runs. Eight of these runs would already have been run; the experimenter needs to run the remaining half.

The experimenter, however, can answer all the questions instead by adding only four more runs. These runs are selected in the following manner: take the four rows of Table 5.2.1 that have '–1' in the 'X_1' column and switch the '–' sign under X_1 to '+' to obtain the four runs given in Table 5.2.2. This is called a foldover on X_1, choosing the subset of runs with $X_1 = -1$. It maybe noted that this choice of 4 runs is not unique, and that if the initial design suggested that $X_1 = -1$ were a desirable level, we would have chosen to experiment at the other four treatment combinations that were omitted from the initial design.

Table 5.2.2. Foldover on X_1 of the 2^4–1 design of Table 5.2.1

Run	X_1	X_2	X_3	X_4
9	1	–1	–1	–1
10	1	1	–1	1
11	1	–1	1	1
12	1	1	1	–1

These four runs together with the eight runs in Table 5.2.1 give the required design in 12 runs.

It is easily verified that no two-factor interaction columns are alike. This means that no two-factor interaction involving X_1 is aliased with any other two-factor interaction. Thus, the design is resolution V, which is not always the case when constructing these types of (3/4) foldover designs.

The runs in Table 5.2.3 give a design with 12 runs, which can estimate all the two-factor interactions involving X_1 free of aliasing with any other two-factor interaction.

Table 5.2.3. A twelve-run design based on the 2^{4-1} (also showing all two-factor interaction columns)

Run	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3	X_2X_4	X_3X_4
1	–1	–1	–1	–1	1	1	1	1	1	1
2	1	–1	–1	1	–1	–1	1	1	–1	–1
3	–1	1	–1	1	–1	1	–1	–1	1	–1
4	1	1	–1	–1	1	–1	–1	–1	–1	1
5	–1	–1	1	1	1	–1	–1	–1	–1	1
6	1	–1	1	–1	–1	1	–1	–1	1	–1
7	–1	1	1	–1	–1	–1	1	1	–1	–1
8	1	1	1	1	1	1	1	1	1	1
9	1	–1	–1	–1	–1	–1	–1	1	1	1
10	1	1	–1	1	1	–1	1	–1	1	–1
11	1	–1	1	1	–1	1	1	–1	–1	1
12	1	1	1	–1	1	1	–1	1	–1	–1

The problem can be viewed more general than the one explained. The problem may be to obtain $\frac{p}{2^k}$ of 2^n (say $\frac{3}{4}(2^n)$ or $\frac{3}{8}$ or $\frac{5}{8}$ or $\frac{7}{8}$ of 2^n, wherever feasible).

Here the 2-symbol OAs of strength two can be useful. The method is as follows:

Start with

1. $\mathbf{H}_u = \mathbf{H}_t \otimes \mathbf{H}_{2^{n-k}}$; $2^{n-k} \geq 2$; $t \geq p$ (an arbitrary scalar).
2. Delete the column of all ones.
3. Choose p subsets out of t-sets in $\mathbf{H}_u$ (which p depends on the contrasts of interest).
4. Choose any n-columns from (2) above (with a care that the treatment combinations are not repeated or repeat is minimum).
5. The resulting design is $\frac{p}{2^k}$ of 2^n.

Note : If the number of runs required is an Hadamard number then as far as possible get hold of

Hadamard matrix of that order; it will avoid the problem of deleting rows.

5.3 Variance Estimation from a Large Scale Complex Survey Data

OAs and MOAs have found interesting application as a Balanced Repeated Replication (BRR) for variance estimation of a non-linear statistic from a large scale complex survey data. Gupta and Nigam (1987) established a linkage between MOAs of strength two and the BRR as a consequence of which it is possible to make unequal number of primary selections from each stratum in a multi stage stratified sampling design for variance estimation of a non-linear statistic like regression and correlation coefficients or birth and death rates, etc. For estimating the variance of nonlinear statistics in stratified sampling designs, the BRR method has received special attention in the literature, although other procedures like linearization (Taylor's series expansion method), Jackknife repeated replications and the Bootstrap method are also available in the literature. BRR method involves forming replications by choosing one of the units selected from each stratum to form a replication. Each of the replications provides an estimate of the non-linear statistic. The procedure is repeated many times to get more stable estimator. The second moment of the estimates obtained from the repeated replications provides the variance estimate.

For two primary selections per stratum, sampling with equal or unequal probabilities and with replacement, McCarthy (1966, 1969) proposed BRR method that involves forming half-samples by randomly selecting one primary sampling unit from the two units in each stratum and showed that using the columns of Plackett and Burman (1946) plans in two symbols (0 and 1) for re-sampling from each of the stratum, there is no loss in efficiency for linear statistics. This is illustrated below with the help of an example.

Example 5.3.1 Consider that the sample design consists of a simple random sample with replacement of size n_h=2 selected from a stratum h with population size N_h, for h = 1, ... , L(=4). Further let $N_1 = 5; N_2 = 6; N_3 = 8; N_4 = 6$; $N=N_1 + N_2 + N_3 + N_4 = 25$. The values of the characteristic under study for the units selected from the 4 strata are respectively:

Stratum	Observations on selected units	Stratum mean = $\bar{y}_h$	$W_h = N_h/N$	$W_h \bar{y}_h$
1	45, 66	55.5	0.20	11.10
2	33, 44	38.5	0.24	9.24
3	35, 42	38.5	0.32	12.32
4	73, 82	77.5	0.24	18.60
$\bar{y}_{st} = \sum_{h=1}^{4} W_h \bar{y}_h$ (an unbiased estimator of population mean)				51.26

It is well known that an unbiased estimator of the variance of $\bar{y}_{st}$ is

$$\hat{v}(\bar{y}_{st}) = \Sigma_h P_n^2 n_h^{-1} \Sigma_i (y_{hi} - \bar{y}_h)^2,$$

where $P_h^2 = W_h^2 / (n_h - 1)$ and $\bar{y}_h = \Sigma_i y_{hi} / n_h$.

For the example, $\hat{v}(\bar{y}_{st}) = 8.57$

Now to obtain an estimate of variance of the estimated mean, we use BRR method. We make use of a Plackett and Burman Plan for 2^4 in 8 runs. In each stratum, let us use the following mapping: $0 \rightarrow y_{h1}$ and $1 \rightarrow y_{h2}$, h = 1,2,3,4 . The choice of mapping, however, is to be made randomly in each stratum:

0	0	0	0	45	33	35	73
0	0	1	1	45	33	42	82
0	1	0	1	45	44	35	82
0	1	1	0	45	44	42	73
1	0	0	1	66	33	35	82
1	0	1	0	66	33	42	73
1	1	0	0	66	44	35	73
1	1	1	1	66	44	42	82

Let y_{hi_j} denote the observation in the j^{th} replication from the h^{th} stratum, j = 1, 2, ..., 8; h = 1, 2, 3, 4 and y_{hi_j} is either of y_{h1} or y_{h2}. For the j^{th} replication, obtain $\bar{y}_j = \Sigma_h P_h y_{hi_j}$. Here $P_h = W_h$ as $n_h = 2$. The various computations involved are shown below:

← Repeated Samples →

	R1	R2	R3	R4	R5	R6	R7	R8
$P_1 y_{1i_j}$	9.00	9.00	9.00	9.00	13.20	13.20	13.20	13.20
$P_2 y_{2i_j}$	7.92	7.92	10.56	10.56	7.92	7.92	10.56	10.56
$P_3 y_{3i_j}$	11.20	13.44	11.20	13.44	11.20	13.44	11.20	13.44
$P_4 y_{4i_j}$	17.52	19.68	19.68	17.52	19.68	17.52	17.52	19.68
$\bar{y}_j = \Sigma_h P_h y_{hi_j}$	45.64	50.04	50.44	50.52	52.00	52.08	52.48	56.88
$(\bar{y}_j - \bar{y}_{BRR})$	–5.62	–1.22	–0.82	–0.74	0.74	0.82	1.22	5.62
$(\bar{y}_j - \bar{y}_{BRR})^2$	31.58	1.49	0.67	0.55	0.55	0.67	1.49	31.58

Now $$\bar{y}_{BRR} = \frac{1}{8}\sum_{j=1}^{8} \bar{y}_j = 51.26$$

and $$\hat{v}(\bar{y}_{BRR}) = \frac{1}{8}\left(\bar{y}_j - \bar{y}_{BRR}\right)^2 = 8.57.$$

Therefore, the results of BRR give unbiased estimator for population mean with same estimate of variance as the usual one.

The set of replications that achieve the full precision is called a *Balanced Set* and the method, therefore, is termed as *BRR*. Since the method consists of selecting one of the two units, the number of units selected for each replication is exactly one half the total sample size. Hence the nomenclature Balanced Half sample method is commonly used to describe the McCarthy's case of BRR. The Plackett and Burman plans are nothing but Hadamard matrices or orthogonal arrays of strength two in two symbols.

Gurney and Jewett (1975) extended the method of balanced half samples to a case when number of primary selections per stratum is p, a prime or a prime power. The set of BRR in this case is once again an OA of strength two in p symbols. Gupta and Nigam (1987) established a linkage between OAs and MOAs of strength two with the BRR used for variance estimation. This linkage enabled to extend the method of BRR to arbitrary number of primary selections per stratum designs. The algebra of the use of mixed orthogonal arrays in variance estimation is given in the sequel.

To simplify exposition attention will be confined to the estimation of a population mean from a stratified random sample. Extension to more complicated situations can be handled by using linear approximations. We suppose the sample design to consist of a simple random sample with replacement of size $n_h \geq 2$ selected from a stratum h with population size N_h, for h = 1, ... , L. The measurement on the i^{th} member of stratum h will be denoted by y_{hi}, for i=1, ..., n_h, so that an unbiased estimator of the population mean is

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \sum_{i=1}^{n_h} y_{hj/n_h} ,$$

where $W_h = N_h/N$, $N=(N_1+ ... +N_L)$.

An unbiased estimator of the variance of $\bar{y}_{st}$ is given by

$$\hat{v}\left(\bar{y}_{st}\right) = \Sigma_h P_h^2 n_h^{-1} \Sigma_i \left(y_{hi} - \bar{y}_h\right)^2 ,$$

where $P_h^2 = W_h^2 / \left(n_h - 1\right)$ and $\bar{y}_h = \Sigma_i y_{hi/n_h}$.

The same estimator of var ($\bar{y}_{st}$) may be calculated by the use of balanced subsamples with one observation per stratum constructed by the use of mixed orthogonal arrays of strength two. In effect, these mixed orthogonal arrays of strength two define balanced subsamples, so that the existence of such an array implies the existence of balanced subsamples with the requisite properties.

To see this define a set of R subsamples with one observation per stratum. The single observation in stratum h for sample j will be denoted by $y_{i_{jh}}$, where i_{jh} is one of 1, ..., n_h. We can equally write $y_{i_{jh}} = \Sigma_i \delta\left(i,\ i_{jh}\right) y_{hi}$, where $\delta(k,\ l) = 1$ if k = l and $\delta(k,\ l) = 0$ otherwise. For the j^{th} subsample let $\bar{y}_j = \Sigma_h P_h y_{i_{jh}}$. Then the average of these terms may be written as

$$\bar{y} = R^{-1} \sum_{h=1}^{L} P_h \sum_{i=1}^{n_h} y_{hi} \sum_{j=1}^{R} \delta\left(i,\ i_{jh}\right),$$

where $\Sigma_j \delta\left(i,\ i_{jh}\right)$ represents the number of times that unit i from stratum h occurs over all R subsamples. If this is constant for all i within h, $\Sigma_j \delta\left(i,\ i_{jh}\right) = \mu_h$, then, as $\Sigma_j \Sigma_i \delta\left(i,\ i_{jh}\right) = R$, we have that $R = \mu_h n_h$ or $\mu_h = R n_h^{-1}$, whence $\bar{y} = \Sigma_h P_h \bar{y}_h$.

Consider now

$$\sum_{j=1}^{R}\left(\bar{y}_j-\bar{y}\right)^2=\sum_{j=1}^{R}\bar{y}_j^2-R\bar{y}^2= S+T-R\bar{y}^2$$

with $S=\sum_{j=1}^{R}\sum_{h=1}^{L}P_h^2\left\{\sum_{i=1}^{n_h}\delta\left(i,i_{jh}\right)y_{hi}\right\}^2$,

$$T=\sum_{j=1}^{R}\sum_{h\neq h'}^{L}P_hP_{h'}\sum_{i=1}^{n_h}\sum_{i'=1}^{n_{h'}}y_{hi}y_{h'i'}\delta\left(i,i_{jh}\right)\ \ \delta\left(i',i_{jh'}\right).$$

As only one unit is selected from each stratum in each subsample, it is readily deducted that $S=R\sum_h P_h^2\sum_i y_{hi}^2 n_h^{-1}$. In summation T, the term $\sum_j \delta(i,\ i_{jh})\ \delta(i',\ i_{jh'})$ represents the number of times that unit i from stratum h appear in same subsample as unit i′ from stratum h′. If this is constant, say $\mu_{hh'}$ for all pairs (i, i′) from strata (h, h′), then, as

$$\sum_{j=1}^{R}\sum_{i=1}^{n_h}\sum_{i'=1}^{n_{h'}}\delta\left(i,i_{jh}\right)\ \ \delta\left(i',i_{jh'}\right)=R=n_hn_{h'}\mu_{hh'}\ ,$$

we have $\mu_{hh'}=R/\left(n_hn_{h'}\right)$. This allows us to write

$$T=\ R\sum_{h\neq h'}P_hP_{h'}\bar{y}_h\bar{y}_{h'}\ ,$$

whence S + T = $R\sum_{h=1}^{L}P_h^2n_h^{-1}\sum_{i=1}^{n_h}\left(y_{hi}-\bar{y}_h\right)^2+R\bar{y}^2$.

Thus $R^{-1}\sum_{j=1}^{R}(\bar{y}_j-\bar{y})^2=\sum_h P_n^2n_h^{-1}\sum_i\left(y_{hi}-\bar{y}_h\right)^2$ = $\hat{v}(\bar{y}_{st})$ as required.

We now illustrate the procedure through an example.

Example 5.3.2 Consider a stratified design involving 50 strata with 6 primary selections from one stratum, 3 primary selections within each of the next 12 strata and 2 primary selections within each of the remaining 37 strata. The observed values are $y_{11}, y_{12} \ldots y_{16}$ in stratum 1, y_{h1}, y_{h2}, y_{h3}, h = 2, 3, ..., 13, in stratum two to thirteen, and y_{r1}, y_{r2}, r = 14, 15, ..., 50, for the remaining strata. Consider the following set of 72 balanced sub-samples row-wise, A, formed with one observation per stratum:

$$\mathbf{A}=\begin{bmatrix}\mathbf{B} & \mathbf{C}\\ \mathbf{B} & \mathbf{D}\end{bmatrix}$$

where **B** and **C** are 36×14 and 36×36 matrices, respectively. Here **D** is obtained from **C** by making the following transformation $0\leftrightarrow 1$. The matrices **B** and **C** = [$\mathbf{C}_1$ $\mathbf{C}_2$] are given in Tables 5.1(a), (b), (c), respectively. The 72 balanced sub-samples form a MOA (72, 50, 6 × 3^{12} × 2^{37}, 2).

Table 5.3.1(a). Matrix **B**

0	0	0	0	1	1	0	0	1	0	2	2	0	0
0	1	1	1	2	2	1	1	2	1	0	0	1	0
0	2	2	2	0	0	2	2	0	2	1	1	2	0
0	0	0	0	0	2	0	2	0	2	0	0	1	1
0	1	1	1	1	0	1	0	1	0	1	1	2	1
0	2	2	2	2	1	2	1	2	1	2	2	0	1
1	0	0	1	0	0	2	1	2	0	0	1	0	0
1	1	1	2	1	1	0	2	0	1	1	2	1	0
1	2	2	0	2	2	1	0	1	2	2	0	2	0
1	0	0	2	2	0	1	0	0	1	1	0	0	1
1	1	1	0	0	1	2	1	1	2	2	1	1	1
1	2	2	1	1	2	0	2	2	0	0	2	2	1
2	0	0	2	2	0	0	1	1	2	0	2	2	0
2	1	1	0	0	1	1	2	2	0	1	0	0	0
2	2	2	1	1	2	2	0	0	1	2	1	1	0
2	0	0	2	1	2	1	2	2	2	2	1	0	1
2	1	1	0	2	0	2	0	0	0	0	2	1	1
2	2	2	1	0	1	0	1	1	1	1	0	2	1
3	0	0	0	0	2	2	0	2	1	1	2	2	0
3	1	1	1	1	0	0	1	0	2	2	0	0	0
3	2	2	2	2	1	1	2	1	0	0	1	1	0
3	0	0	1	2	1	2	2	0	0	2	0	2	1
3	1	1	2	0	2	0	0	1	1	0	1	0	1
3	2	2	0	1	0	1	1	2	2	1	2	1	1
4	0	0	1	2	1	0	0	2	2	1	1	1	0
4	1	1	2	0	2	1	1	0	0	2	2	2	0
4	2	2	0	1	0	2	2	1	1	0	0	0	0
4	0	0	1	0	0	1	2	1	1	2	2	1	1
4	1	1	2	1	1	2	0	2	2	0	0	2	1
4	2	2	0	2	2	0	1	0	0	1	1	0	1
5	0	0	2	1	2	2	1	1	0	1	0	1	0
5	1	1	0	2	0	0	2	2	1	2	1	2	0
5	2	2	1	0	1	1	0	0	2	0	2	0	0
5	0	0	0	1	1	1	1	0	1	0	1	2	1
5	1	1	1	2	2	2	2	1	2	1	2	0	1
5	2	2	2	0	0	0	0	2	0	2	0	1	1

Table 5.3.1(b). Matrix C_1

1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	1	1	1	0	0	0	1	1	1	1	1	0	1	1
1	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1	0	1
1	1	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1	0
1	0	1	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1
1	0	0	1	0	0	1	0	1	1	1	0	0	0	1	1	1	1
1	1	0	0	1	0	0	1	0	1	1	1	0	0	0	1	1	1
1	1	1	0	0	1	0	0	1	0	1	1	1	0	0	0	1	1
1	0	1	1	0	0	1	0	0	1	0	1	1	1	0	0	0	1
1	1	0	1	1	0	0	1	0	0	1	0	1	1	1	0	0	0
1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	1	0	0
1	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	1	0
1	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	1
1	0	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1
1	0	0	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1
1	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0	1	0
1	1	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0	1
1	0	1	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0
1	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0	1	0
1	1	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0	1
1	1	1	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0
1	1	1	1	0	0	1	0	0	0	0	1	0	1	0	1	1	0
1	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0	1	1
1	1	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0	1
1	1	1	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0
1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	1	0	1
1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	1	0
1	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	1
1	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0
1	0	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0
1	0	0	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0
1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0	1	0
1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0	1
1	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0
1	0	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0
1	1	0	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1

Table 5.3.1(c). Matrix C_2

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0	1	0
1	1	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0	1
1	1	1	0	0	1	0	0	0	0	1	0	1	0	1	1	0	0
0	1	1	1	0	0	1	0	0	0	0	1	0	1	0	1	1	0
1	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0	1	1
1	1	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0	1
1	1	1	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0
1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	1	0	1
1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	1	0
0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	1
0	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0
0	0	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0
1	0	0	0	1	1	1	1	1	0	1	1	1	0	0	1	0	0
1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0	1	0
1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0	1
0	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0
1	0	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0
0	1	0	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1
0	0	1	0	1	1	1	0	0	0	1	1	1	1	1	0	1	1
1	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1	0	1
0	1	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1	0
0	0	1	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1
1	0	0	1	0	0	1	0	1	1	1	0	0	0	1	1	1	1
1	1	0	0	1	0	0	1	0	1	1	1	0	0	0	1	1	1
0	1	1	0	0	1	0	0	1	0	1	1	1	0	0	0	1	1
1	0	1	1	0	0	1	0	0	1	0	1	1	1	0	0	0	1
0	1	0	1	1	0	0	1	0	0	1	0	1	1	1	0	0	0
1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	1	0	0
0	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	1	0
0	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	1
0	0	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1
0	0	0	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1
1	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0	1	0
0	1	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0	1
0	0	1	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0

Thus, a clear relationship exists between the MOAs of strength two and the set of BRR. The method of BRR is applicable whenever a MOA of strength two exists. However, the MOAs of strength two do not always exist for all the combinations of symbols. This puts a severe restriction on the application of BRR method in real sample survey situations. Kish and Frankel (1970) introduced a method called grouping method to overcome the obstacle of non-availability of a balanced set in the case of arbitrary number of primary selections. The method is to divide the sample sizes within each stratum at random into two equal subsamples, and then to form balanced repeated replications treating each subsample as if it is one member of a sample of size 2. This method is commonly called as grouped balanced half sample method or grouped balanced repeated replication. In fact, the repeated replications based on this method are partially balanced. Wu (1991) has shown that the grouping method is very inefficient to estimate the variance of a non-linear statistic. Wu (1991) discouraged the use of proportional frequency plans in the BRR as these result in efficiency loss in estimating the variance of a non-linear statistic. Wu advocated the use of near orthogonal arrays in which most of the columns are orthogonal.

Dhandapani *et al.* (1996) advocated the use of orthogonal main effect plans with unequal frequencies in general and with proportional frequencies in particular. Further, they have also shown that the variance estimate using proportional frequency plans is asymptotically consistent for a non-linear statistic that is expressible as a general smooth function of population means.

5.4 Computer Experiments

Definition 5.4.1 A k-dimensional Latin Hypercube Design (LHD) of n points, is a set of n points $x_i = (x_{i1}, x_{i2}, ..., x_{ik}) \in \{0, 1, ..., n-1\}^k$ such that for each dimension j, all x_{ij} are distinct. An LHD is called maximin when the separation distance $\min_{i \neq j} d(\mathbf{x}_i, \mathbf{x}_j)$ is maximal among all LHDs of given size n, where d is a certain distance measure. Such maximin LHDs are very useful as designs for computer experiments.

A 10-Dimenional L2 Latin hypercube design of 16 points; d = 1.296148140 and D = 378

0	2	5	0	8	6	2	6	11	4
1	8	7	7	12	15	11	3	9	15
2	14	2	9	14	1	7	7	4	6
3	15	9	3	3	14	10	12	5	1
4	6	3	12	5	7	15	5	15	2
5	12	13	14	1	5	9	4	3	11
6	7	10	2	7	2	13	13	14	14
7	9	15	4	13	10	0	9	0	8
8	11	0	10	4	9	1	10	13	13
9	3	1	6	9	11	12	15	2	9
10	1	8	15	11	0	3	11	6	12
11	0	14	5	0	13	6	8	8	10
12	4	12	8	15	4	14	0	7	7
13	13	6	1	2	3	8	1	10	5
14	10	11	11	10	8	5	14	12	0
15	5	4	13	6	12	4	2	1	3

It may be seen from Definition 5.3 and the example that an OA (n, m, n, 1) is a Latin hypercube. Latin hypercubes have been frequently used in conducting computer experiments. Applying an orthogonal Latin hypercube design to a computer experiment benefits the data analysis in two ways. First, it retains the orthogonality of traditional experimental designs. The estimates of linear effects of all factors are uncorrelated not only with each other, but also with the estimates of all quadratic effects and bilinear interactions. Second, it can facilitate nonparametric fitting procedures, because one can select good space-filling designs within the class of orthogonal Latin hypercubes according to selection criteria.

ACKNOWLEDGEMENTS

The author is grateful to the Indian Society of Agricultural Statistics for giving an opportunity to deliver the Technical Address during the 61st Annual Conference of the Society held at Birsa Agricultural University, Ranchi, during 30 November to 02 December 2007. The author is thankful to Dr. Aloke Dey, Dr. Rajender Parsad, Dr. Kishore Sinha and Dr. Rahul Mukerjee for their help and suggestions that have led to considerable improvements in the manuscript. The author is also grateful to Dr. Bikas Sinha and Dr. Sugata Adhikari for helpful discussions on irregular fractional factorials.

REFERENCES

Addelman, S. and Kempthorne, O. (1961). Some main effect plans and orthogonal arrays of strength two. *Ann. Math. Statist.,* **32**, 1167-1176.

Agrawal, V. and Dey, A. (1982). A note on orthogonal main effect plans for asymmetric factorials. *Sankhya,* **B44**, 278-282.

Ai, Mingyao, Fang, Kai-Tai and He, Shuyuan (2007). $E(\chi^2)$-optimal mixed-level supersaturated designs. *J. Statist. Plann. Inf.,* **137**, 306-316.

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

Booth, K.H.V. and Cox, D.R. (1962). Some systematic supersaturated designs. *Technometrics,* **4**, 489-495.

Bulutoglu, Dursun A. and Cheng, C.S. (2004). Construction of $E(s^2)$-optimal supersaturated designs. *Ann. Statist.,* **32**, 1662-1678.

Butler, N., Mead, R., Eskridge, K.M. and Gilmour, S.G. (2001). A general method of constructing $E(s^2)$-optimal supersaturated designs. *J. Roy. Statist. Soc.,* **B63**, 621-632.

Chacko, A. and Dey, A. (1981). Some orthogonal main effect plans for asymmetrical factorials. *Sankhya,* **B43**, 384-391.

Cheng, C.S. (1989). Some orthogonal main effect plans for asymmetrical factorials. *Technometrics,* **31**, 475-477.

Cheng, C.S. (1997). Optimal supersaturated designs. *Statist. Sinica,* **74**, 929-939.

Das, Ashish, Dey, A., Chan, Ling-Yau and Chatterjee, Kashinath (2007). On $E(s^2)$-optimal supersaturated designs. *J. Statist. Plann. Inf.* (Communicated) [Paper seen with the courtesy of authors].

Dey, A. (1985). *Orthogonal Fractional Factorial Designs*. Wiley Eastern Limited, New Delhi.

Dey, A. and Agrawal, V. (1985). Orthogonal fractional plans for asymmetrical factorials derivable from orthogonal arrays. *Sankhya,* **B47**, 56-66.

Dey, A. and Midha, Chand K. (1996). Construction of some asymmetrical orthogonal arrays. *Statist. Probab. Lett.*, **28**, 211-217.

Dey, A. and Mukerjee, R. (1999). *Fractional Factorial Plans*. John Wiley, New York.

Dhandapani, A., Gupta, V.K. and Nigam, A.K. (1996). Variance estimation using proportional frequency plans. *J. Ind. Soc. Agril. Statist.,* **49**, 267-276.

Fang, K.T., Ge, G., Liu, M.Q. and Qin, H. (2004). Combinatorial constructions for optimal supersaturated designs. *Discrete Mathematics,* **279**, 191-202.

Fang, K.T., Lin, D.K.J. and Liu, M.Q. (2003). Optimal mixed-level supersaturated design. *Metrika,* **58**, 279-291.

Fang, K.T., Lin, D.K.J. and Ma, C.X. (2000). On the construction of multi-level supersaturated designs. *J. Statist. Plann. Inf.,* **86**, 239-252.

Gupta, V.K., Dhandapani, A. and Parsad, Rajender (2007). Monograph on Hadamard matrices. IASRI, New Delhi.

Gupta, V.K., Parsad, Rajender, Bhar, L.M. and Kole, B. (2007). Supersaturated designs for asymmetrical factorial experiments. *J. Statist. Theo. Practice,* **2(1)**, 95-108.

Gupta, V.K., Nigam, A.K. and Dey, A. (1982). Orthogonal main effect plans for asymmetrical factorials. *Technometrics,* **24(2)**, 135-137.

Gupta, V.K. and Nigam, A.K. (1987). Mixed orthogonal arrays for variance estimation with unequal numbers of primary selections per stratum. *Biometrika,* **74**, 735-742.

Gurney, M. and Jewett, R.S. (1975). Constructing orthogonal replications for variance estimation. *J. Amer. Statist. Assoc.,* **71**, 819-821.

Hedayat, A.S., Pu, K. and Stufken, J. (1992). On the construction of asymmetrical orthogonal arrays. *Ann. Statist*., **20**, 2142-2152.

Hedayat, A.S., Sloane, N.J.A. and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer, New York.

Kennedy, M.C. and O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika,* **87**, 1–13.

Kish, L. and Frankel, R.M. (1970). Balanced repeated replications for standard errors. *J. Amer. Statist. Assoc.,* **65**, 1071-1094.

Kohli, P. (2006). A study on supersaturated designs. Unpublished M.Sc. dissertation, IARI, New Delhi.

Lejeune, M.A. (2003). A coordinate-columnwise exchange algorithm for construction of supersaturated, saturated and non-saturated experimental designs. *Amer. Jour. Math. Manag. Sci.,* **23(1-2)**, 109-142.

Lu, X. and Sun, Y. (2001). Supersaturated design with more than two levels. *Chinese Ann. Math.*, **B22**, 183-194.

McCarthy, P.J. (1966). Replication: An approach to the analysis of data from complex surveys, Vital and Health Statistics, Series 2, No. 14, Washington, D.C., US Department of Health Education and Welfare, National Centre for Health Statistics.

McCarthy, P.J. (1969). Psuedo replication: Half samples. *Rev. Int. Statist. Inst.,* **37**, 239-264.

Mukerjee, R., Qian, Z. and Wu, C.F.J. (2007). On the existence of nested orthogonal arrays. Paper seen with the courtesy of the authors.

Nguyen, N.K. (1996). An algorithmic approach to constructing supersaturated designs. *Technometrics,* **38**, 69-73.

Plackett, R.L. and Burman, J.P. (1946). Designs of optimum multi-factorial experiment. *Biometrika,* **33**, 305-325.

Qian, Z., Seepersad, C., Joseph, R., Allen, J. and Wu, C.F.J. (2006). Building surrogate models with detailed and approximate simulations. *ASME Journal of Mechanical Design,* **128**, 668-677.

Qian, Z. and Wu, C.F.J. (2006). Bayesian Hierarchical modeling for integrating low accuracy and high accuracy experiments. *Technometrics* (communicated).

Rao, C.R. (1947). Factorial arrangements derivable from combinatorial arrangements of arrays. *J. Roy. Statist. Soc. Suppl.,* **9**, 118-139.

Rao, C.R. (1973). Some combinatorial problems of arrays and applications in design of experiments. In : *A Survey of Combinatorial Theory (J.N. Srivastava et al. Eds)*, *North Holland, Amsterdam*, 349-359.

Rees, C.S., Wilson, A.G., Hamada, M., Martz, H.F. and Ryan, K.J. (2004). Integrated analysis of computer and physical experiments. *Technometrics,* **46**, 153-164.

Ryan, Kenneth J. and Bulutoglu, Dursun A. (2007). $E(s^2)$-optimal supersaturated designs with good minimax properties. *J. Statist. Plann. Inf.,* **137**, 2250-2262.

Sinha, K., Vellaisamy, P. and Sinha, N. (2008). Kronecker sum of binary orthogonal arrays. *Utilitas Mathematica* (to appear) [Paper seen with the courtesy of the authors].

Tang, B. and Wu, C.F.J. (1997). A method for constructing supersaturated designs and its $E(s^2)$optimality. *Canad. J. Statist.*, **25(2)**, 191-201.

Wang, P.C. (1990). On the constructions of some orthogonal main effect plans. *Sankhya,* **B 52**, 319-323.

Wu, C.F.J. and Hamada, M. (2000). *Experiments: Planning, Analysis and Parameter Design Optimization.* John Wiley, New York.

Wu, C.F.J. (1991). Balanced repeated replications based on mixed orthogonal arrays. *Biometrika,* **78**, 181-188.

www.iasri.res.in/design/webhadamard/

www.iasri.res.in/design/supersaturated/

Yamada, S. and Lin, D.K.J. (1999). Three level supersaturated designs. *Statist. Probab. Lett.,* **45**, 31-39.