Further Contributions to Balanced Arrays of Strength Four

D.V. Chopra
Department of Mathematics and Statistics
Wichita State University
Wichita, KS 67260-0033, USA
dharam.chopra@wichita.edu

Richard M. Low
Department of Mathematics
San Jose State University
San Jose, CA 95192, USA
low@math.sjsu.edu

R. Dios
Department of Mathematics
New Jersey Institute of Technology
Newark, NJ 07102-1982, USA
dios@adm.njit.edu

Abstract

In this paper, we present (by using Cauchy-Schwarz inequalities) some new results amongst the parameters of balanced arrays (B-arrays) with two symbols and having strength four which are necessary for the existence of such balanced arrays. We then discuss and illustrate their use and applications.

1 Introduction and Preliminaries

For the sake of completeness, we first present some basic definitions and concepts. An *array* T with m rows (also called *constraints* or *factors* in design of experiments), N columns (also called *runs* or *treatment-combinations*), and with two symbols (also called *levels*) is merely a matrix T of size $(m \times N)$ with two elements (say, 0 and 1). If $\underline{\alpha}$ is a column vector of T, then $\lambda(\underline{\alpha})$, $P(\underline{\alpha})$, and $w(\underline{\alpha})$ denote respectively the number of times $\underline{\alpha}$ occurs in T, the vector obtained by permuting the elements of $\underline{\alpha}$, and the *weight* of the vector $\underline{\alpha}$ (ie. the number of 1s in it). It is quite obvious that $w(\underline{\alpha})$ is invariant under a permutation of the elements of $\underline{\alpha}$. Under various combinatorial structures, these arrays assume great importance in combinatorics and statistical design of experiments. Here, we will confine ourselves to some such constraints leading us to the following definition:

Definition. A matrix T $(m \times N)$ with two symbols (say, 0 and 1) is called a *balanced array* (B-array) of *strength* t $(t \leq m)$ if in every $(t \times N)$ submatrix T^* of T (clearly, there are $\binom{m}{t}$ such submatrices T^*), every $(t \times 1)$ vector $\underline{\alpha}$ of weight i (clearly, $0 \leq w(\underline{\alpha}) \leq t$) occurs with the same frequency μ_i (say, $0 \leq i \leq t$).

Remark. The vector $\underline{\mu}' = (\mu_0, \mu_1, \ldots, \mu_t)$ is called the *index set* of the B-array T, and obviously $N = \sum_{i=0}^{t} \binom{t}{i} \mu_i$. Thus, N is known once we are given $\underline{\mu}'$. The above definition can be easily extended to B-arrays with more than two levels.

Definition. If $\mu_i = \mu$ for each i in a B-array T, then T is called an *orthogonal array* (O-array).

Note that $N = 2^t \mu$ for an O-array of strength t and that O-arrays form a subset of B-arrays.

These arrays have been extensively used in statistical design of experiments which are widely used in almost all areas of scientific investigations such as medicine, technology, industry, agriculture, etc. O-arrays (a special case of B-arrays) have been used in cryptography, coding theory, computer science, information theory, and in the famous Taguchi techniques on quality control widely used in industry. Bose [2] applied O-arrays to information theory to point out the connections between the problems of experimental designs and information theory. B-arrays are also related to other combinatorial structures. For example, the incidence matrix of a balanced incomplete block design (BIBD) is a B-array of strength two. Saha, et. al. [11] have pointed out the relationship of B-arrays of strength two with rectangular designs, group divisible designs, nested balanced incomplete block designs, etc. Thus, B-arrays are not only useful in numerous scientific investigations but also are of great use in the study of other combinatorial structures. To gain further insight into the importance and usefulness of B-arrays and O-arrays to statistical design of experiments and to combinatorics, the interested reader may consult the list of references (which, by no means is an exhaustive one) at the end of this paper, and also further references mentioned therein.

The problem of the existence of a B-array for a given m ($\geq t+1$) and a given $\underline{\mu}'$ is obviously a very non-trivial problem. In this paper, we present a set of new inequalities involving the parameters m and $\underline{\mu}'$. For a B-array T to exist, it is necessary that each and every inequality presented here be satisfied. Also, we can obtain the maximum number m of constraints for a given $\underline{\mu}'$, which is an important problem from the point of view of design of experiments and combinatorics. Such problems for O-arrays have been discussed, among others, by Bose and Bush [1], C.R. Rao [9, 10], Seiden and Zemach [12], and for B-arrays by Chopra, Bsharat, Dios, and/or Low [3, 4, 5], Rafter and Seiden [8], Saha et. al. [11], Yamamoto et. al. [14], etc. In this paper, we obtain some such inequalities for B-arrays with $t = 4$, $\underline{\mu}' = (\mu_0, \mu_1, \mu_2, \mu_3, \mu_4)$, $m \geq 4$, and illustrate their use and applications.

2 Main Results with Discussion

The following results can be easily derived.

Lemma 1. *A B-array T of strength $t = 4$, with index set $\underline{\mu}'$ and $m = 4$, always exists.*

Lemma 2. *A B-array T of strength $t = 4$ is also of lower strength k (≤ 4).*

Remark. Considered as an array of strength k ($\leq t = 4$), the elements of the index set of T are merely a linear combination of the elements of $\underline{\mu}'$. If $A(j,k)$ is the jth element of T (when considered as an array of strength k, where $0 \leq j \leq k$), then

$$\begin{aligned} A(j,k) &= \sum_{i=0}^{t-k} \binom{t-k}{i} \mu_{i+j} \\ &= \sum_{i=0}^{4-k} \binom{4-k}{i} \mu_{i+j}, \quad \text{where } t \text{ is set equal to 4.} \end{aligned} \tag{2.1}$$

It is clear that $A(j,4) = \mu_j$ and $A(j,0) = N$.

Lemma 3. *If a B-array T with index set $\underline{\mu}'$ does not exist for an m (say, $m = m^*$), then it does not exist for all $m > m^*$.*

The next result relates the parameters of the array T with the moments of the weights of the column vectors of T, and can be easily established by counting in two ways (through columns and rows) the total number of vectors of weight k ($0 \leq k \leq 4$).

Lemma 4. *Consider a B-array T with m rows, of strength $t = 4$, and of index set $\underline{\mu}'$. Let x_j ($j = 0, 1, 2, \ldots, m$) be the number of columns of weight* *j in T. Let* $L_k = \sum_{j=0}^{m} j^k x_j$*, with* $(0 \le k \le 4)$*. Then, the following results are true:*

$$\begin{aligned} L_0 &= \sum_{j=0}^{m} x_j = N, \\ L_1 &= \sum j x_j = m_1 A(1,1), \\ L_2 &= L_1 + m_2 A(2,2), \\ L_3 &= 3L_2 - 2L_1 + m_3 A(3,3), \\ L_4 &= 6L_3 - 11L_2 + 6L_1 + m_4 \mu_4, \end{aligned} \tag{2.2}$$

where $m_r = m(m-1)(m-2)\cdots(m-r+1)$.

It is obvious, from (2.2), that each of the $\sum j^k x_j$ is a polynomial function in m for a given $\underline{\mu}'$.

Next, we state some new results on the existence of B-arrays of strength $t = 4$ and prove some of them, while providing outlines and sketches of others. In what follows, we use the symbols a and b to denote $\overline{j}$ and $\overline{j^2}$ respectively (ie. $a = \overline{j} = \frac{\sum j x_j}{N} = \frac{L_1}{N}$, and $b = \overline{j^2} = \frac{\sum j^2 x_j}{N} = \frac{L_2}{N}$).

Theorem 1. *For a B-array T with t = 4 and m rows to exist, the following condition must be satisfied:*

$$N^2 L_4 - 2N L_1 L_3 + L_1^2 L_2 \ge 0. \tag{2.3}$$

Proof. (Outline). Consider the inequality $\sum j^2 (j-a)^2 x_j \ge 0$. By expanding the left-hand side, one obtains (2.3). □

Theorem 2. *Let T be a B-array with* $m \ge 4$ *and* $\underline{\mu}' = (\mu_0, \mu_1, \mu_2, \mu_3, \mu_4)$*. Then, the following must be satisfied for T to exist:*

$$N L_4 + N L_2 + 2 L_1 L_2 \ge L_1^2 + L_2^2 + 2N L_3. \tag{2.4}$$

Proof. (Outline). In order to derive (2.4), consider the inequality

$$\sum [(j-a) - (j^2 - b)]^2 x_j \ge 0.$$

Expanding the left-hand side, one obtains

$$\begin{aligned} \sum [(j-a) - (j^2-b)]^2 x_j &= \sum [(j^2 - 2aj + a^2) + (j^4 - 2j^2 b + b^2) \\ &\quad - 2(j^3 - bj - aj^2 + ab)] x_j \\ &= (L_2 - 2aL_1 + a^2 N) + (L_4 - 2bL_2 + b^2 N) \\ &\quad - 2(L_3 - bL_1 - aL_2 + abN). \end{aligned}$$

Substituting $\frac{L_1}{N}$ for a and $\frac{L_2}{N}$ for b, we obtain (2.4) after some simplication. □

We now recall the Cauchy-Schwarz Inequality, which we will use later.

Cauchy-Schwarz Inequality. If $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are sequences of real numbers, then

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right).$$

Theorem 3. *For a B-array T with parameters $\underline{\mu}' = (\mu_0, \mu_1, \mu_2, \mu_3, \mu_4)$ and m to exist, we must have the following:*

$$N(N^3L_4 - 4N^2L_1L_3 + 6NL_1^2L_2 - 3L_1^4)L_4 \geq (N^2L_4 - 2NL_1L_3 + L_1^2L_2)^2, \tag{2.5}$$

$$NL_2(N^3L_4 - 4N^2L_1L_3 + 6NL_1^2L_2 - 3L_1^4) \geq (N^2L_3 - 2NL_1L_2 + L_1^3)^2, \tag{2.6}$$

$$(N^3L_4 - 4N^2L_1L_3 + 6NL_1^2L_2 - 3L_1^4) \geq (NL_2 - L_1^2)^2. \tag{2.7}$$

Proof. We will use the Cauchy-Swartz Inequality on $(j-a)^2\sqrt{x_j}$ and $j^2\sqrt{x_j}$ to obtain (2.5). Thus, we get $(\sum(j-a)^4x_j)(\sum j^4x_j) \geq [\sum j^2(j-a)^2x_j]^2$. As $a = \frac{L_1}{N}$, we see that

$$\begin{aligned}(\textstyle\sum(j-a)^4x_j)(\sum j^4x_j) &= L_4(L_4 - 4L_3a + 6L_2a^2 - 4L_1a^3 + a^4N)\\ &= L_4\left(L_4 - \frac{4L_1L_3}{N} + 6\frac{L_1^2L_2}{N^2} - 4\frac{L_1^4}{N^3} + \frac{L_1^4}{N^3}\right)\\ &= \frac{L_4}{N^3}(N^3L_4 - 4N^2L_1L_3 + 6NL_1^2L_2 - 3L_1^4).\end{aligned}$$

A similar calculation shows that

$$[\textstyle\sum j^2(j-a)^2x_j]^2 = \frac{(N^2L_4 - 2NL_1L_3 + L_1^2L_2)^2}{N^4}.$$

Hence, (2.5) is established.

In order to derive (2.6), we use the Cauchy-Schwarz Inequality on $(j-a)^2\sqrt{x_j}$ and $j\sqrt{x_j}$. Similarly, (2.7) can be established by applying the Cauchy-Schwarz Inequality on $(j-a)^2\sqrt{x_j}$ and $\sqrt{x_j}$. The results are obtained after some straightforward algebraic manipulations. □

Theorem 4. *For a B-array T with parameters $\underline{\mu}' = (\mu_0, \mu_1, \mu_2, \mu_3, \mu_4)$ and* *m to exist, we must have the following:*

$$\begin{aligned} N(N^3L_4 - 4N^2L_1L_3 + 6NL_1^2L_2 - 3L_1^4)(N^2L_2 - 2NL_1L_2 + L_2^2) \\ \geq (N^3L_4 - 2N^2L_2L_3 + NL_2^3 - 2N^2L_1L_3 + 4NL_1L_2^2 - L_1^2L_2^2 \\ + NL_1^2L_2 - 2L_1^3L_2)^2, \end{aligned} \tag{2.8}$$

$$\begin{aligned} N(N^3L_4 - 4N^2L_1L_3 + 6NL_1^2L_2 - 3L_1^4)(L_4 - 2L_3 + L_2) \\ \geq (N^2L_4 - 2NL_1L_3 + L_1^2L_2 - N^2L_3 + 2NL_1L_2 - L_1^3)^2. \end{aligned} \tag{2.9}$$

Proof. (Outline). In order to derive (2.8), we use the Cauchy-Schwarz Inequality with $(j-a)^2\sqrt{x_j}$ and $(j-b)^2\sqrt{x_j}$ to obtain $(\sum(j-a)^4x_j)(\sum(j-b)^4x_j) \geq [\sum(j-a)^2(j-b)^2x_j]^2$. After simplifying the left-hand side and right-hand side of this inequality and setting $a = \frac{L_1}{N}$, $b = \frac{L_2}{N}$, we obtain (2.8).

To establish (2.9), we use the Cauchy-Schwarz Inequality with $(j-a)^2\sqrt{x_j}$ and $(j^2-j)\sqrt{x_j}$ to get $(\sum(j-a)^4x_j)(\sum(j^2-j)^2x_j) \geq [\sum(j-a)^2(j^2-j)x_j]^2$. The left-hand side of this inequality simplifies to

$$\left(L_4 - \frac{4L_1L_3}{N} + \frac{6L_1^2L_2}{N^2} - \frac{3L_1^4}{N^3}\right)(L_4 - 2L_3 + L_2).$$

The right-hand side of the inequality simplifies to

$$\begin{aligned} &[\textstyle\sum(j^4 - 2j^3a + j^2a^2 - j^3 + 2j^2a - a^2j)x_j]^2 \\ &= \left(L_4 - \frac{2L_1L_3}{N} + \frac{L_1^2L_2}{N^2} - L_3 + \frac{2L_1L_2}{N} - \frac{L_1^3}{N^2}\right)^2. \end{aligned}$$

After multiplying both sides of the inequality by N^4 and simplifying further, one obtains (2.9). □

3 Some Comments and Illustrative Examples

To obtain the least value of m for which a B-array T with a given $\underline{\mu}'$ could possibly exist, we prepared a computer program. It is obvious that each of the results given here was reduced to a polynomial inequality in m. Starting with $m = 4$, let us suppose the inequality was first contradicted at $m = m^* + 1$. Then, the maximum value of m for which this array could possibly exist is $m = m^*$. We worked with scores of values of $\underline{\mu}'$ and found there was no single inequality superior to all the rest, for each $\underline{\mu}'$.

Examples.

- Take $\underline{\mu}' = (6,4,1,0,0)$. It was found that $m \leq 8$ for each of (2.3), (2.4), and (2.8) while other conditions gave $m \leq 13, 19$, and exceedingly large values. Taking $\underline{\mu}' = (9,8,7,7,5)$, we found $m \leq 9$ using (2.8) which is the least one and $m \leq 18$ using (2.3), $m \leq 34$ using (2.4). Taking $\underline{\mu}' = (1,1,6,3,1)$, we found $m \leq 6$ (the least one) using (2.4) while all the other inequalities gave us very large m values.

Next, we give some examples to show how these results compare with earlier ones found within the literature.

- Take $\underline{\mu}' = (4,4,4,4,3)$. Using (2.4) of Chopra [3], it was found that $m \leq 32$. Using the results in this paper, we found $m \leq 11$ using (2.8), $m \leq 12$ using each of (2.5), (2.6) and (2.9). This is a considerable improvement over the earlier bound.

- Take $\underline{\mu}' = (1,1,2,4,1)$. Using (2.6) of Chopra and Bsharat [4], it was found that $m \leq 6$. If we use (2.3) of this paper, we obtain $m \leq 5$ which is an improvement over the earlier result.

The above discussion and illustrations are merely presented to indicate that no single inequality amongst the parameters of a B-array may provide us with the least upper on the number of constraints m, for all the B-arrays with given values of $\underline{\mu}'$.

References

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

[2] R.C. Bose, On some connections between the design of experiments and information theory, *Bull. Internat. Statist. Inst.* 38 (1961), 257-271.

[3] D.V. Chopra, A note on an upper bound for the constraints of balanced arrays with strength t, *Commun. Statist. Theor. Meth.* 12(15) (1983), 1755-1759.

[4] D.V. Chopra and M. Bsharat, Contributions to balanced arrays of strength four on two symbols, *Congr. Numer.* 157 (2002), 183-189.

[5] D.V. Chopra, R.M. Low and R. Dios, On the maximum number of constraints for some balanced arrays, *Congr. Numer.* 190 (2008), 5-11.

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

[7] J.Q. Longyear, Arrays of strength t on two symbols, *J. Statist. Plann. Inf.* 10 (1984), 227-239.

[8] J.A. Rafter and E. Seidon, Contributions to the theory and construction of balanced arrays, *Ann. Statist.* 2 (1974), 1256-1273.

[9] C.R. Rao, Hypercubes of strength d leading to confounded designs in factorial experiments, *Bull. Calcutta Math. Soc.* 38 (1946), 67-78.

[10] C.R. Rao, Some combinatorial problems of arrays and applications to design of experiments, *A Survey of Combinatorial Theory* (edited by J.N. Srivastava, et. al.) (1973), North-Holland Publishing Co., 349-359.

[11] G.M. Saha, R. Mukerjee and S. Kageyama, Bounds on the number of constraints for balanced arrays of strength t, *J. Statist. Plann. Inf.* 18 (1988), 255-265.

[12] E. Seiden and R. Zemach, On orthogonal arrays, *Ann. Math. Statist.* 27 (1966), 1355-1370.

[13] W.D. Wallis, *Combinatorial Designs*, Second Edition. (2007), Marcel Dekker Inc., New York.

[14] S. Yamamoto, M. Kuwada and R. Yuan, On the maximum number of constraints for s-symbol balanced arrays of strength t, *Commun. Statist. Theory Meth.* 14 (1985), 2447-2456.