Balanced Arrays and Fractional Factorial Designs

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

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

Introduction and Preliminaries

For the sake of completeness and ease of reference, we state some basic concepts and definitions.

Definition. An array with m rows (constraints), N columns (runs, treatment-combinations), and with s symbols (say, $0, 1, \ldots, s-1$; also called levels) is merely a matrix T of size $m \times N$ with elements from the s symbol set.

Definition. An array T of size $(m \times N)$ and with two symbols (0 and 1) is called a balanced array of strength t $(t \leq m)$ if in every t-rowed submatrix T^* of T (clearly, there are $\binom{m}{t}$ such submatrices), every $(t \times 1)$ vector $\underline{\alpha}$ of weight i $(0 \leq i \leq t$; the weight of a vector is the number of nonzero elements) and its permutation $P(\underline{\alpha})$ appears with the same frequency, (say) μ_i. The vector $\underline{\mu}' = (\mu_0, \mu_1, \ldots, \mu_t)$ is called the index set of T, and clearly, $N = \sum_{i=0}^{t} \binom{t}{i} \mu_i$.

It is quite obvious that orthogonal arrays (O-arrays) form a subset of B-arrays under the condition that $\mu_i = \mu$ (for each i). The incidence matrix of a balanced incomplete block (BIB) design is also a B-array with $t = 2$. The relationship of balanced arrays of strength two with rectangular designs, group divisible designs and nested balanced incomplete block designs has been described in [20]. Orthogonal arrays with two symbols have been used to construct fractional factorial designs for 2^m series, but the main problem in using O-arrays is that these arrays do not exist for each N (the number of treatment-combinations). For example, for $t = 4$, the number of columns N in the O-array must be a multiple of 2^4 (ie. 16). Thus, O-arrays with $t = 4$ can only exist if $N = 16\mu$ (μ is a positive integer ≥ 1). To overcome this drawback, the concept of B-arrays was introduced (at the suggestion of C.R. Rao) by Chakravarti [3]. Relaxing the combinatorial constraint on O-arrays gave rise to B-arrays. B-arrays with two levels and with different values of the strength t have been extensively used to construct fractional factorial designs of 2^m series having different resolutions. For example, B-arrays with $t = 4$, $s = 2$ have been extensively used to construct optimal balanced fractional factorial designs of resolution V (ie. designs which allow us to estimate all the effects up to and including two-factor interactions under the assumption that higher order interactions are negligible). To gain further insight into the importance of B-arrays to design of experiments and their relationship to other combinatorial structures, the interested reader may consult the list of references (by no means an extensive one) at the end of this paper, and also further references listed therein.

To construct B-arrays (O-arrays) of strength t $(t > m)$ for a given $\underline{\mu}'(\mu)$ and m is a very nontrivial and complex problem. This problem becomes all the more challenging if our interest is to construct such arrays with the maximum possible value of m for a given $\underline{\mu}'$ (ie. N is known). The resolution of this problem will amount to accommodating more factors (for a given N) in the language of design theory, which means savings in cost of running the experiment. Such problems for B-arrays and O-arrays have been investigated, among others, by Bose/Bush, Rao [2, 16], Seiden/Zemach [19], and by Chopra/Low and Dios, Rafter/Seiden, and Saha et. al. in [8, 9, 10, 11, 12, 18], etc.

In this paper, we confine ourselves to B-arrays with $t = 5$. Balanced fractional factorial designs obtained by using B-arrays with $t = 5$, under certain conditions, would give rise to resolution six designs (ie. designs which allow us to estimate all the effects up to and including two-factor interactions in the presence of three-factor interactions when higher-order interactions are negligible).

Main Results on B-arrays with $t = 5$ with Applications

The following results are either quite obvious or can be easily established.

Lemma 2.1. A B-array with $m = t$ always exists.

Lemma 2.2. A B-array T of strength t with $\underline{\mu}' = (\mu_0, \mu_1, \ldots, \mu_t)$ is also of strength t', where $0 < t' \leq t$.

Note that when T is considered as an array of strength t' $(0 < t' \leq t)$, the jth element $(0 \leq j \leq t)$ of its parameter-vector is given by $A_{j,t'} = \sum_{i=0}^{t-t'} \binom{t-t'}{i} \mu_{i+j}$, with the convention that $\binom{a}{b} = 1$ if $a = b = 0$. It is quite obvious that $A_{j,t'}$ is merely a linear combination of the elements $\mu_j, \mu_{j+1}, \ldots, \mu_{j+(t-t')}$, for $0 \leq j \leq t'$.

Lemma 2.3. Let x_j $(0 \leq j \leq m)$ denote the frequency of the columns of weight j in a B-array T $(m \times N)$ of strength $t = 5$ with $\underline{\mu}' = (\mu_0, \mu_1, \ldots, \mu_5)$. For T to exist, the following conditions must be satisfied:

$$\begin{aligned} L_k &= \sum_{j=0}^{m} j^k x_j = N, \quad \text{if } k = 0; \\ &= \sum_{r=1}^{k} a_r m_r A_{r,r}, \quad \text{for } 1 \leq k \leq 5, \end{aligned} \tag{0.1}$$

where $m_r = m(m-1)\cdots(m-r+1)$, values of a_r are known, and $A_{r,r}$ are also known if $\underline{\mu}'$ is given.

Clearly, (0.1) expresses the L_ks as polynomials in m. We provide next, for ease of computations, the values of a_r for various values of k. The elements of the vector $(a_1, a_2, \ldots, a_k)$ for $1 \leq k \leq 5$, are respectively: (1), $(1, 1)$, $(1, 3, 1)$, $(1, 7, 6, 1)$, and $(1, 15, 25, 10, 1)$.

Theorem 2.1. Let x_j $(0 \leq j \leq m)$ be the number of columns of weight j in a B-array T $(m \times N)$ of strength five and with index set $\underline{\mu}' = (\mu_0, \mu_1, \ldots, \mu_5)$. In order for T to exist, the following inequalities must be satisfied:

(a) $L_1 L_3 \geq L_2^2$.

(b) $L_1 L_3 L_5 + 2 L_2 L_3 L_4 \geq L_1 L_4^2 + L_2^2 L_5 + L_3^3$.

Proof. Consider the following matrix of moments

$$W_5 = \begin{pmatrix} L_1 & L_2 & L_3 \\ L_2 & L_3 & L_4 \\ L_3 & L_4 & L_5 \end{pmatrix}.$$

It is a symmetric matrix and is non-negative definite (n.n.d.), which can be seen by observing the non-negative definiteness of the quadratic form $\sum_{j=0}^{m}(\alpha_1 j^{1/2} + \alpha_2 j^{3/2} + \alpha_3 j^{5/2})^2 x_j$ in variables α_1, α_2, and α_3. Since W_5 is n.n.d., all its leading principal minors have determinant greater than or equal to zero. The leading principal minors are

$$\begin{pmatrix} L_1 & L_2 \\ L_2 & L_3 \end{pmatrix}$$

and W_5. We obtain the two inequalities, respectively, by setting the determinants of these minors ≥ 0.

$\square$

Next, we give some illustrative examples. In order to accomplish this, we prepared a computer program to test the inequalities in Theorem 2.1 for values of $\underline{\mu}'$ and $m \geq 6$. For a given $\underline{\mu}'$ and m, if any inequality is contradicted, then the B-array for that $\underline{\mu}'$ and m does not exist. Clearly, this would also allow us to obtain the max(m) for a given $\underline{\mu}'$. We compare both inequalities and also these against the ones published earlier in [9, 11].

Example 1. For $\underline{\mu}' = (2, 3, 3, 3, 3, 2)$, inequality (2.1a) gives $m \leq 48$ while (2.1b) gives $m \leq 10$. Thus in this case, inequality (2.1b) is better. For $\underline{\mu}' = (1, 1, 3, 1, 1, 2)$, we obtain $m \leq 6$ using (2.1b), but (2.1a) did not get contradicted even for $m \geq 100$. For $\underline{\mu}' = (2, 0, 0, 0, 1, 2)$, we find (2.1b) is not contradicted even for $m \geq 100$ but (2.1a) gives $m \leq 7$. Hence, no one result is superior in each case. Now, we compare the present results with those published in [9, 11].

Example 2. The max(m), as published in [9], for the arrays $(1, 4, 1, 1, 1, 1)$, $(1, 1, 1, 7, 4, 3)$, and $(2, 5, 7, 1, 1, 1)$ are found to be 6, 8, and 7, respectively; where as the max(m) for the above arrays (using 2.1b) are found to be 5, 7, and 7, respectively. For the array with $\underline{\mu}' = (1, 1, 1, 1, 2, 1)$, the max$(m)$ is 7, as published in [11], but (2.1b) gives us $m \leq 6$.

In conclusion, we see that none of these conditions is uniformly better for every B-array. However, we observed some results may fare better in comparison to others. For example, (2.1b) gave us sharper bounds on m as compared to (2.1a) in the majority of the B-arrays selected by us. These inequalities are very useful in the sense that they remove from consideration, all $\underline{\mu}'$ and m for which any one is contradicted.

REFERENCES (RÉFERENCES)

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

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

[3] I.M. Chakravarti, Fractional replication in asymmetrical factorial designs and partially balanced arrays., *Sankhya* 17 (1956), 143-164.

[4] I.M. Chakravarti, On some methods of construction of partially balanced arrays., *Ann. Math. Statist.* 32 (1961), 1181-1185.

[5] C.S. Cheng, Optimality of some weighing and 2^n fractional designs., *Ann. Statist.* 8 (1980), 436-446.

[6] D.V. Chopra and J.N. Srivastava, Optimal balanced 2^7 fractional factorial designs of resolution V, with $N \leq 42$., *Ann. Inst. Stat. Math.* 25, #3 (1973), 587-604.

[7] D.V. Chopra and J.N. Srivastava, Optimal balanced 2^7 fractional factorial designs of resolution V, $43 \leq N \leq 48$., *Sankhya: The Indian Journal of Statistics* 37, Series B, Part 4, (1975), 429-447.

[8] 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.

[9] D.V. Chopra and M. Bsharat, On the constraints of some combinatorial arrays., *J. Combin. Math. Combin. Comput.* 62 (2007), 159-164.

[10] D.V. Chopra and R.M. Low, Contributions to balanced arrays of strength t with applications., *Frontiers in Applied and Computational Mathematics*, edited by D. Blackmore, et. al., World Scientific, New Jersey, (2008).

[11] D.V. Chopra, M. Bsharat and G.P. Mehta, Some new results on balanced arrays of strength five., *J. Combin. Math. Combin. Comput.* 66 (2008), 59-64.

[12] D.V. Chopra, R.M. Low and R. Dios, Contributions to strength six balanced arrays using Holder and Minkowski inequalities., *J. Combin. Math. Combin. Comput.* 72 (2010), 93-100.

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

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

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

[16] C.R. Rao, Factorial experiments derivable from combinatorial arrangements of arrays., *J. Roy. Statist. Soc. Suppl.* 9 (1947) 128-139.

[17] 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., North-Holland Publishing Co., (1973), 349-359.

[18] 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.

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

[20] K. Sinha, V. Dhar, G.M. Saha and S. Kageyama, Balanced arrays of strength two from block designs., *Combin. Designs* 10 (2002), 303-312.

RÉSUMÉ (ABSTRACT)

In this paper, we define balanced arrays (B-arrays) discuss their relationships with other combinatorial structures in design of experiments, and describe their use in the construction of fractional factorial designs. In particular, we obtain some necessary existence conditions for B-arrays with $s = 2$ *symbols and of strength* $t = 5$*. As a consequence, we obtain, for a given* $\underline{\mu}' = (\mu_0, \mu_1, \ldots, \mu_5)$*, the maximum value of the number of constraints* m*.*