BU-689-M

NEW COMBINATORIAL DESIGNS AND THEIR APPLICATIONS TO GROUP TESTING*

D. Raghavarao, H. Pesotan and W. T. Federer

Abstract

A class of designs with property C(t) are introduced for the first time and their applications in group testing of experiments are studied.

1. *Introduction*.

Let us consider the problem of classifying each of n given units into one of two disjoint categories called satisfactory and unsatisfactory (or, simply, good and bad or defective). The characteristic feature of group testing is that any number of units, say x, can be tested simultaneously, but the information obtained from a single test on x units, without any chance of error, is that either (i) all the x units are good, or (ii) at least one of the x units tested is bad, but it is unknown how many and which ones are bad. The problem is to devise a suitable method of classifying all the n units into good or bad categories with the least number of trials.

The first application of group testing in the literature was made by Dorfman [2] in pooling blood samples in order to classify each one of a large group of people as to whether or not they have a particular disease. Sobel and his co-workers [5], [7], [8], [9], [10]

*Research supported by grant no. A8766 from the National Research Council of Canada and the NIH grant no. GM-05900.

have devised various sequential procedures to classify the units and established the optimality of their results for large n. Lindström [3], [4] was interested in a slightly modified problem, in which each trial determines the exact number of defectives, and provided optimal procedures in a set-theoretic frame.

In Section 2 we introduce a new class of combinatorial designs, which we call designs with completeness property on t symbols, written as property $C(t)$ and study them in some detail. We then use them in group testing of experiments in Section 5. For terminology in combinatorics of design of experiments, we refer to Raghavarao [6].

2. *A New Combinatorial Design.*

Let S be a set of v symbols $1, 2, \ldots, v$ and let $B_1, B_2, \ldots, B_b$ be non-empty proper subsets of S for $i = 1, 2, \ldots, b$. The design D is the collection of subsets $B_1, B_2, \ldots, B_b$ along with the set of symbols S. We now define the following:

DEFINITION 2.1. *The design* D *is said to have the completeness property on* t *symbols, shortly written as the* property $C(t)$, *if for every* t *distinct symbols* $\theta_1, \theta_2, \ldots, \theta_t \in S$,

$$\bigcup_{j \in T} B_j = S - \{\theta_1, \theta_2, \ldots, \theta_t\} \tag{2.1}$$

where $T = \{j \mid \theta_i \notin B_j \text{ for } i = 1, 2, \ldots, t\}$.

The balanced incomplete block design (BIB design)

$$(0, 1, 3); (1, 2, 4); (2, 3, 5); (3, 4, 6); (4, 5, 0);$$
$$(5, 6, 1); (6, 0, 2) \tag{2.2}$$

with parameters

$$v = 7 = b \ , \quad r = k = 3 \ , \quad \lambda = 1 \tag{2.3}$$

has the property C(2) . For example, let us consider the symbols 0, 5 . The sets in which none of the symbols 0, 5 occur are (1, 2, 4), (3, 4, 6) and the union of these two sets is {1, 2, 3, 4, 6} . Similarly, if we consider the symbols 2, 6 ; the sets in which none of the symbols 2, 6 occur are (0, 1, 3), (4, 5, 0) whose union is {0, 1, 3, 4, 5} .

Trivially a C(t) design exists for $1 \le t \le v$, with $b=v$. In fact, the design with $B_i = \{i\}$, for $i = 1, 2, \ldots, v$ is a C(t) design for $1 \le t \le v$. We call such a design a trivial C(t) design.

The class of designs to be looked into to obtain C(t) designs are not necessarily the BIB designs alone. Any kind of design may possess the property C(t) . The result regarding the C(2) property in BIB designs is contained in the following:

THEOREM 2.1. *A* BIB *design with parameters* v, b, r, k, λ *possesses the* property C(2) *if and only if*

$$r - 2\lambda > 0 \ . \tag{2.4}$$

Proof. Let θ_1, θ_2 be any 2 symbols of the design. It is well known that there are $b - 2r + \lambda$ sets of the design which do not contain the symbols either θ_1 or θ_2 . The sufficiency part of the proof will be completed if we show that of the remaining $v - 2$ symbols each occurs at least once among those $b - 2r + \lambda$ sets. If ϕ is a symbol of the design other than θ_1 , θ_2 , it can occur at most 2λ times in the sets where there is at least one of θ_1 , θ_2 and hence it must occur at least $r - 2\lambda$ (> 0) times in the sets where there is none of θ_1 , θ_2 . Conversely, since every symbol other than θ_1 , θ_2 occurs at least once in the sets where there is none of θ_1, or θ_2 and as this number should be at least $r - \lambda$, it follows that $r - 2\lambda \geq 1$ or equivalently (2.4) proving the necessity part of the theorem.

From Theorem 2.1 it follows, for example, that the BIB design with parameters $v = 7 = b$, $r = 4 = k$, $\lambda = 2$ does not have the property C(2).

In searching for designs with the C(t) property in the known classes of designs, the following theorem will be helpful.

THEOREM 2.2. *If a design* D *with sets* $S_1, S_2, \ldots, S_b$ *and* S *as the set of symbols has the* property C(t), *then in the complimentary design* D^* *formed from the sets* $S_1^*, S_2^*, \ldots, S_b^*$, *the number of times every* t-*plet of symbols* $(\theta_1, \theta_2, \ldots, \theta_t)$ *occur, denoted by* $\lambda_{\theta_1\theta_2\ldots\theta_t}$ *is a positive integer, where* $S_i^* = S - S_i$, *for* $i = 1, 2, \ldots, b$.

Proof. When the design D has property C(t), there exist sets, say, $S_{i_1}, S_{i_2}, \ldots, S_{i_x}$, where a given t-plet of symbols $(\theta_1, \theta_2, \ldots, \theta_t)$ do not occur, while all the other symbols occur at least once. Then in D^*, the blocks $S_{i_1}^*, S_{i_2}^*, \ldots, S_{i_x}^*$ will each contain the symbols $\theta_1, \theta_2, \ldots, \theta_t$ and hence $\lambda_{\theta_1\theta_2\ldots\theta_t} = x\ (> 0)$.

The condition stated in the theorem is only necessary, but not sufficient. The BIB design with parameters $v = 7 = b$, $r = 4 = k$, $\lambda = 2$ has its complimentary design in which every symbol occurs at least once satisfying the condition of the theorem, but does not possess the C(2) property as indicated after Theorem 2.1.

It is well known that $C[k, \ell, \delta, v]$ configurations are also $C[k, \ell', \delta', v]$ configurations. Analogous to this result we have the following:

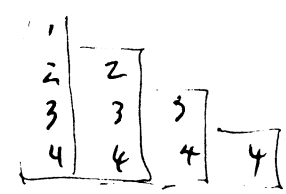
THEOREM 2.3. *If a design* D *has the* property C(t) , *then it has the* property C(t-1) *for* $2 \le t \le v$.

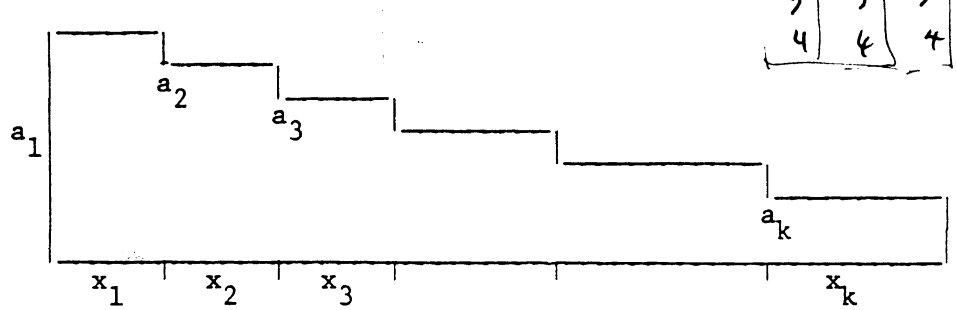
Proof. Let $(\theta_1, \theta_2, \ldots, \theta_t)$ be any t-plet. Among the sets where at least one of the symbols $\theta_1, \theta_2, \ldots, \theta_t$ occur; for each θ_i $(i = 1, 2, \ldots, t)$ there exist at least one set in which θ_i occurs without $\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_t$. For, otherwise for the t-plet $(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_t, \phi)$ with $\phi \neq \theta_i$ the property C(t) for the design D will be violated. Now the sets in which none of $(\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_t)$ occur all the other $v-t+1$ symbols occur proving that D has property C(t-1) .

In view of the group testing situations for which these designs are proposed, we need the designs with property C(t) for which $b<v$. We discuss these results in the next 2 sections.

3. *Designs with* $b<v$ *Having* property C(1) .

Any v can be written in the form $a_1x_1 + a_2x_2 + \ldots + a_kx_k$ where $k = 1, 2, \ldots$ and a_i, x_i are positive integers. Without loss of generality, we assume $a_1 > a_2 > \ldots > a_k$. The numbers can then be written in the form of a staircase:





We then form b blocks where $b = a_1 + x_1 + x_2 + \ldots + x_k$ by writing the symbols in the a_1 rows and the $x_1 + x_2 + \ldots + x_k$ columns.

Such designs will have exactly 2 replications for each symbol and have various cardinalities for the sets constituting the design. A moment's consideration into the above construction indicates that such designs have the property C(1) .

Let us illustrate our construction method for $v=19$. Since $19 = 4\times4 + 3\times1$, we write the 19 symbols in the staircase array

$$\begin{array}{ccccc} 1 & 2 & 3 & 4 & \\ 5 & 6 & 7 & 8 & 19 \\ 9 & 10 & 11 & 12 & 18 \\ 13 & 14 & 15 & 16 & 17 \end{array} \tag{3.1}$$

We now form 9 sets for the design by writing the sets formed from the rows and columns of the array (3.1) to get

$$\begin{array}{l} (1, 2, 3, 4);\ (5, 6, 7, 8, 19);\ (9, 10, 11, 12, 18); \\ (13, 14, 15, 16, 17);\ (1, 5, 9, 13);\ (2, 6, 10, 14); \\ (3, 7, 11, 15);\ (4, 8, 12, 16);\ (17, 18, 19)\ . \end{array} \tag{3.2}$$

The design (3.2) can be easily verified to have the property C(1) .

It is interesting to note that as the partitioning of v as $a_1x_1 + a_2x_2 + \ldots + a_kx_k$ is not unique and different partitionings give different numbers of blocks, it is desirable to consider the partitioning for a given v which minimizes $b = a_1 + x_1 + x_2 + \ldots + x_k$.

We now study the existence of C(1) designs with $b<v$ and the asymptotic property for b/v as $v \to \infty$. Their results are given in the following theorem:

THEOREM 3.1. *Designs with* property C(1) *and* $b<v$ *exist for all* $v\geq 6$. *Furthermore,*

$$\lim_{v\to\infty} \frac{b}{v} = 0 . \qquad (3.3)$$

Proof. Let $m^2 < v \leq (m+1)^2$, for $m = 0, 1, \ldots$. We distinguish 2 cases: case (i) $v = m^2 + a$, where $1\leq a\leq m$ and case (ii) $v = m^2 + m + b$, where $1\leq b\leq m+1$. In case (i) clearly $v = m\times m + a\times 1$ and from our earlier consideration a design with property C(1) can be constructed in $b = m + (m+1) = 2m + 1$ sets. In case (ii) we have $v = (m+1)\times m + b\times 1$ and we can construct a design with property C(1) in $b = (m+1) + (m+1) = 2(m+1)$ sets. Now $2m + 1 < m^2 + a$, where $1\leq a\leq m$ for all $m\geq 3$ and $m=2$, $a=2$. Again $2m + 2 < (m+1)m + b$ where $1\leq b\leq m+1$ for all $m\geq 3$. These considerations imply that $b<v$ for all $v\geq 6$. Now

$$\frac{b}{v} = \frac{2m+i}{v} = o(\frac{1}{m}) , \qquad (3.4)$$

where $i = 1$ or 2 depending on whether v belongs to case (i) or (ii), and the assertion (3.3) follows.

The designs with property C(1) constructed by the above staircase method will not always give the smallest b and this follows from the following theorem:

THEOREM 3.2. *If* D_i $(i = 1, 2)$ *are designs with* property C(1) *on* v_i *symbols in* b_i *sets, then there exists a design* $\tilde{D}$ *with* property C(1) *on* $v_1 v_2$ *symbols in* $(b_1 + b_2)$ *sets.*

Proof. Let S_i be the set of v_i symbols and let $S = S_1 X S_2$, where 'X' is the Cartesian product of sets. Let $B_{1i}, B_{2i}, \ldots, B_{b_i i}$ be the sets of the design D_i . Consider the $b_1 + b_2$ sets $S_1 X B_{j2}$ and

$B_{\ell 1} X S_2$ for $j = 1, 2, \ldots, b_2$ and $\ell = 1, 2, \ldots, b_1$ constituting the design $\tilde{D}$ on $v_1 v_2$ symbols of $S_1 X S_2$. It can easily be verified that $\tilde{D}$ has property C(1) .

From Theorem 3.1 we have a design with property C(1) on 8 symbols in 6 sets. From this design, using Theorem 3.2, we can construct a design with property C(1) on 64 symbols in 12 sets. The design given in Theorem 3.1 on 64 symbols with property C(1) has 16 sets, while Theorem 3.2 leads us to a design with only 12 sets. Consequently, for $64^2 = 4096$ symbols from Theorem 3.2, we can construct a design with 24 sets, while the design constructable from Theorem 3.1 has 128 sets. Thus we achieve considerable reduction in the number of sets used in the design by using the method of Theorem 3.3.

However, in general it remains an open problem to find the smallest b for designs with property C(1) on v symbols.

4. *Designs with* Property C(2) *with* $b<v$.

We have seen in Section 2 that BIB designs have the property C(2) if and only if $r - 2\lambda > 0$. However, such designs will have $b \geq v$ and are not useful in group testing experimental situations. While searching for designs with property C(2) , the class of designs to be looked at are those Partially Balanced Incomplete Block Designs (PBIB designs) for which $b<v$ and whose complimentary designs have positive λ parameters.

While scanning through the designs given in the Tables of Two-Associate-Class Partially Balanced Designs [1], the authors found that SR41 and T85 have the property C(2) . These are designs in 12 symbols and 9 sets, and in 36 symbols with 28 sets respectively, and both of these designs have the property C(2) .

5. *Applications of Designs with* property C(2) *in Group Testing Experiments.*

Let there be v units in the population and let it be known to the experimenter before hand that there are exactly t units in the population which are defective, while v-t units are good. Further, it is unknown to the experimenter which of the units are defective. Then one can make b tests (or runs) on the v units, where each test is made on the collection of the units belonging to the sets of a design D with property C(t) . If the test gives a negative result, the units involved in the test are all good and if the test gives a positive result, at least one of the units involved in the test is bad. If x tests give negative results in each test and the remaining b-x tests give positive results, the C(t) property of the design guarantees that the units, included in the set union of the sets corresponding to the negative test results, are all good which will be v-t in number, while the other v-t are bad.

As an illustration, let us consider that there are 12 units among which we know that 2 are bad and 10 are good. We indicated in Section 4 that the design SR41 of the Tables [1] has property C(2) . The test number and the units tested in each are as follows:

Test Number	Units Included in the Test
1	1, 2, 3, 4
2	7, 10, 5, 4
3	6, 11, 9, 4
4	1, 7, 6, 8
5	11, 5, 2, 8
6	10, 9, 3, 8
7	1, 11, 10, 12
8	9, 2, 7, 12
9	5, 3, 6, 12

The classification of items and the test numbers indicating negative results are as follows:

Test Number	Defective Items	Test Number	Defective Items
2, 3, 6, 9	1, 2	7, 8, 9	4, 8
2, 3, 5, 8	1, 3	4, 5, 7, 9	4, 9
5, 6, 8, 9	1, 4	4, 5, 8, 9	4, 10
3, 6, 8	1, 5	4, 6, 8, 9	4, 11
2, 5, 6, 8	1, 6	4, 5, 6	4, 12
3, 5, 6, 9	1, 7	1, 6, 7, 8	5, 6
2, 3, 8, 9	1, 8	1, 3, 6, 7	5, 7
2, 5, 9	1, 9	1, 3, 7, 8	5, 8
3, 5, 8, 9	1, 10	1, 4, 7	5, 9
2, 6, 8, 9	1, 11	1, 3, 4, 8	5, 10
2, 3, 5, 6	1, 12	1, 4, 6, 8	5, 11
2, 3, 4, 7	2, 3	1, 3, 4, 6	5, 12
4, 6, 7, 9	2, 4	1, 5, 6, 7	6, 7
3, 4, 6, 7	2, 5	1, 2, 7, 8	6, 8
2, 6, 7, 9	2, 6	1, 2, 5, 7	6, 9
3, 6, 7, 9	2, 7	1, 5, 8	6, 10
2, 3, 7, 9	2, 8	1, 2, 6, 8	6, 11
2, 4, 7, 9	2, 9	1, 2, 5, 6	6, 12
3, 4, 9	2, 10	1, 3, 7, 9	7, 8
2, 4, 6, 9	2, 11	1, 5, 7, 9	7, 9
2, 3, 4, 6	2, 12	1, 3, 5, 9	7, 10
4, 5, 7, 8	3, 4	1, 6, 9	7, 11
3, 4, 7, 8	3, 5	1, 3, 5, 6	7, 12
2, 5, 7, 8	3, 6	1, 2, 7, 9	8, 9
3, 5, 7	3, 7	1, 3, 8, 9	8, 10
2, 3, 7, 9	3, 8	1, 2, 8, 9	8, 11
2, 4, 5, 7	3, 9	1, 2, 3	8, 12
3, 4, 5, 8	3, 10	1, 4, 5, 9	9, 10
2, 4, 8	3, 11	1, 2, 4, 9	9, 11
2, 3, 4, 5	3, 12	1, 2, 4, 5	9, 12
4, 6, 7, 8	4, 5	1, 4, 8, 9	10, 11
5, 6, 7, 8	4, 6	1, 3, 4, 5	10, 12
5, 6, 7, 9	4, 7	1, 2, 4, 6	11, 12

In view of Theorem 3.1, if a large population has exactly 1 bad item, it can be detected in b tests, where b is only a very tiny fraction of v .

The statistical properties of our test procedure and the comparison of our technique to the known procedure are expected to be discussed in a later communication.

Acknowledgement.

The authors wish to thank Professor B. Leo Raktoe for many helpful discussions.

REFERENCES

[1] W.H. Clatworthy, *Tables of Two-Associate-Class Partially Balanced Designs*, NBS Appl. Math. Series 63 (1973).

[2] R. Dorfman, *The Detection of Defective Members of Large Populations*, Ann. Math. Statist. 14 (1943), 436-440.

[3] B. Lindström, *On a Combinatory Detection Problem*, I. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, IX (1964), 195-207.

[4] B. Lindström, *Determining Subsets of Unramified Experiments* (1973), Unpublished paper.

[5] S. Kumar and M. Sobel, *Finding a Single Defective in Binomial Group Testing*, J. Amer. Statist. Assoc. 66 (1971), 824-828.

[6] D. Raghavarao, *Constructions and Combinatorial Problems in Design of Experiments*, Wiley (1971).

[7] M. Sobel, *Group-testing to Classify all Defectives in a Binomial Sample*, a contribution in Information and Decision Processes, R. E. Machol, ed., McGraw-Hill (1960), 127-161.

[8] M. Sobel, *Optimal Group Testing*, Proc. Colloquim on Information Theory Organized by the Bolyai Mathematical Society, Debrecen (Hungary) (1967), 411-488.

[9] M. Sobel and P. A. Groll, *Group Testing to Eliminate Efficiently all Defectives in a Binomial Sample*, Bell System Tech. J. 38 (1959), 1179-1252.

[10] M. Sobel and P. A. Groll, *Binomial Group Testing with an Unknown Proportion of Defectives*, Technometrics 8 (1966), 631-656.

Temple University
Philadelphia, Pennsylvania

University of Guelph
Guelph, Ontario

Cornell University
Ithaca, New York