MINIMAL PAIRWISE BALANCED DESIGNS

R. G. STANTON

Department of Computer Science, University of Manitoba, Winnipeg, Canada, R3T 2N2

An expression involving a "remainder term" is given for the number of blocks in a minimal pairwise balanced design in which the length of the longest block is specified. The allows a simple presentation and unification of a number of earlier results derived by various authors.

1. Introduction

Suppose that we are given a set $V$ made up of $v$ elements 1, 2, 3, ..., $v$. A pairwise balanced design is a collection $F$ of blocks with the property that every pair of elements from $V$ occurs exactly $\lambda$ times among the blocks of $F$. In the rest of this paper, we shall restrict attention to the particular case $\lambda = 1$. We shall also introduce the parameter $k$ to designate the length of the longest block in the family $F$ (this block may not be unique; usually, there will be several blocks of length $k$).

As a simple example, let us look at the case $v = 7$, $k = 4$. There are six non-isomorphic pairwise balanced designs with these parameters, and it is instructive to list them.

(a) Blocks 1234, 1567, 9 pairs; total of 11 blocks.
(b) Blocks 1234, 567, 12 pairs; total of 14 blocks.
(c) Blocks 1234, 156, 257, 367, 6 pairs; total of 10 blocks.
(d) Blocks 1234, 156, 257, 9 pairs; total of 12 blocks.
(e) Blocks 1234, 156, 12 pairs; total of 14 blocks.
(f) Blocks 1234, 15 pairs; total of 16 blocks.

It is clear that the minimal pairwise balance design with $v = 7$, $k = 4$, is the design labelled (c).

In general, we use the symbol $g^{(k)}(1, 2; v)$ to designate the minimum cardinality of any pairwise balanced design on a set of $v$ elements with longest block having length $k$. Thus, we have shown, by exhaustive search, that $g^{(4)}(1, 2; v) = 10$. Of course, the minimal design may not be unique; it is perfectly possible for two non-isomorphic designs to possess the same minimal cardinality.

We shall frequently abbreviate $g^{(k)}(1, 2; v)$ to $g^{(k)}(v)$ or simply, in this paper, to $g$. 
2. Elementary relations

In the minimal design, we let \( b_i \) represent the number of blocks of length \( i \), where \( i < k \). If \( i = k \), we designate one particular block of length \( k \) to be the “longest block”, and we use \( b_k \) to designate the number of other blocks of length \( k \). Thus, the total number of blocks of length \( k \) is \( b_k + 1 \). We often refer to the designated “longest block” as the base block; it plays a very specialized role in the theory.

By counting blocks, and then by counting appearances of pairs within blocks, we immediately obtain two relations.

\[
b_2 + b_3 + b_4 + b_5 + \cdots + b_k = g - 1 \quad (1)
\]

\[
2b_2 + 6b_3 + 12b_4 + 210b_5 + \cdots + k(k - 1)b_k = v(v - 1) - k(k - 1)
= (v - k)(v + k - 1). \quad (2)
\]

To obtain a third relation, we define \( b_{ij} \) to be the number of blocks of length \( i \) that pass through point \( j \) on the base block \( (j = 1, 2, 3, \ldots, k) \). Since every pair containing \( j \) must appear in the set of blocks, we immediately have

\[
\Sigma_i(i - 1)b_{ij} = v - k, \quad (3)
\]

and this result holds for every point \( j \). Hence we may sum over \( j \) and obtain

\[
\Sigma_i\Sigma_j(i - 1)b_{ij} = k(v - k). \quad (4)
\]

This summation is over all blocks of length \( i \) that meet the base block. However, there may be some blocks of length \( i \) that are disjoint from the base block; suppose that the number of these is \( b_{i0} \). Then we may form the sum

\[
\Sigma_i(i - 1)b_{i0} = E, \quad (5)
\]

where the quantity \( E \) (for excess) is certainly nonnegative. Since we know that

\[
b_i = b_{i0} + b_{i1} + b_{i2} + b_{i3} + \cdots + b_{ik}, \quad (6)
\]

we can add equations (4) and (5) to end up with

\[
b_2 + 2b_3 + 3b_4 + 4b_5 + \cdots + (k - 1)b_k = k(v - k) + E. \quad (7)
\]

We now combine equations (1), (2), and (7) in such a way as to eliminate adjacent columns in the equations. For instance, using multipliers 2, 1, -4, would eliminate the terms in \( b_3 \) and \( b_4 \) to leave

\[
2(b_2 + 3b_3 + 6b_4 + \cdots).
\]

We shall multiply the three equations by \( s(s + 1) \), 1, \(-2(s + 1)\), respectively, in order to eliminate those terms involving \( b_{s+1} \) and \( b_{s+2} \). The resulting expression involves the quantity

\[
P = b_s + b_{s+3} + 3(b_{s-1} + b_{s+4}) + 6(b_{s-2} + b_{s+5}) + 10(b_{s-3} + b_{s+6}) + \cdots \quad (8)
\]

It is clear that \( P \) is nonnegative.
The result of combining
\[ s(s + 1)(1) + (2) - 2(s + 1)(7) \]
is the relation
\[ s(s + 1)(g - 1) + (v - k)(v + k - 1) - 2(s + 1)k(v - k) = 2E(s + 1) + 2P. \tag{9} \]
If we solve for \( g \) from Eq. (9), the result is
\[ g = 1 + (v - k)(2sk - v + k + 1)/s(s + 1) + 2E/s + 2P/s(s + 1), \tag{10} \]
where the quantities \( E \) and \( P \) are non-negative. If we drop the terms in \( E \) and \( P \), we obtain a lower bound that was established by Stinson [5] in 1982, using generalized variance techniques.

**Theorem 1** (Stinson). \( g \geq 1 + (v - k)(2sk - v + k + 1)/s(s + 1) \).

This result is true for all values of \( s \); we can easily determine the most effective value for \( s \) by writing \( F(s) = 1 + (v - k)(2sk - v + k + 1)/s(s + 1) \); then we find
\[ F(s) - F(s - 1) = 2(v - k)(v - 1 - sk)/s(s - l)(s + 1). \]
This equation shows that \( F(s) \) is increasing so long as \( sk \) lies below \( (v - 1) \). Hence, to obtain the strongest result from (10), we should assign to \( s \) the value \([v - 1]/k\]; of course, if the quantity \((v - 1)/k\) should happen to be an integer, then both \( F(s) \) and \( F(s - 1) \) are equal.

Now, let us consider the case of a very long block whose length \( k \) lies between \( v/2 \) and \( v \). For \( k \) in this region, we select \( s = 1 \), and thus obtain a result due to Woodall [6].

**Theorem 2** (Woodall). \( If \) \( k \) \( lies \) \( between \) \( v/2 \) \( and \) \( v \), \( then \) \( g \geq 1 + (v - k)(3k - v + 1)/2 \).

We note that the Woodall bound is always an integer. Consequently, Eq. (9) can be applied to give

**Corollary 2.1.** The Woodall bound can only be achieved if \( F = P = 0 \), that is, all blocks meet the long base block, and their lengths are either 2 or 3.

This bound can actually be met by using an easy construction based on 1-factors of the \((v - k)\) points not in the long block; see [4] for details.

However, Eq. (9) gives us more information than simply the Woodall bound and its converse. Suppose that we now let \( k \) lie between \( v/3 \) and \( v/2 \); then we take \( s = 2 \). (We should remark that special techniques may have to be applied when one is at the exact boundary of this region, that is, where \( s \) is changing from 1 to 2 or from 2 to 3.) In this case, the term \( 2E/s \) in (9) becomes \( E \); because \( E \) is
a non-negative integer, we see that $E$ must be zero if the Stinson bound is met. If we write $S$ for the Stinson bound, and require that it be "met" (that is, $g = \lfloor S \rfloor$), then we have

$$g = S + 2P/s(s + 1) = S + (b_2 + b_3)/3,$$

where the second term is less than unity. Consequently, we have

**Theorem 3.** If $k$ lies between $v/3$ and $v/2$, and the Stinson bound is met (in the nearest-integer sense), then $E = 0$, that is, all blocks meet the base block. Furthermore, all of the blocks have lengths 3 or 4, except that there may possibly be one or two rogue blocks (this corresponds to the case $P = 1$ or $P = 2$), and the number of these is given by the relation

$$[S] - S = (b_2 + b_3)/3.$$

There is currently a great deal of work being done for $k$ lying in this region; see, for example [3], the very important work of Rees in [1] and [2], and the various works cited in [1] and [2]. The use of "frames" (cf. [1]) has been of particular significance in discussing the question.

Actually, Theorem 3 is only a special case of a more general result. Suppose that the Stinson bound is actually met, that is, $g = \lfloor S \rfloor$. Then we prove, without any restriction on $k$, that is, for all values of $s \geq 2$,

**Theorem 4.** The Stinson bound can only be met, that is, $g = \lfloor S \rfloor$, if all of the blocks meet the long block.

**Proof.** We suppose that, if possible, the Stinson bound is met, but that there is a block of length $(s + 1) - z$ that does not meet the base block. This block will contribute an amount $(s - z)$ to $E$; however, it also contributes an amount $z(z + 1)/2$ to $P$. There is a certain balancing effect in action here, since small $z$ values make $E$ large and $P$ small, whereas large $z$ values make $P$ large and $E$ small. More precisely, we may write

$$g = S + 2E/s + 2P/s(s + 1),$$

where the contribution of the disjoint block to the "remainder terms" is given by

$$2(s - z)/s + z(z + 1)/s(s + 1) = (z^2 - z(2s + 1) + 2s(s + 1))/s(s + 1).$$

(11)

Now the discrete variable $z$ may range from the value 1, if there is a disjoint block of length $s$, to the value $(s - 1)$, if there is a disjoint block of length 2. The expression (11) is decreasing and reaches its minimum value (in the permissible range for $z$) at $s - 1$; this minimum value is

$$(s^2 + s + 2)/(s^2 + s),$$
Minimal pairwise balanced designs

and it is greater than unity. Consequently, it is not possible to have \( g = \lceil S \rceil \) unless there is no disjoint block, that is, \( E = 0 \), as stated in Theorem 4.

It is an obvious corollary that, if the Stinson bound is met (that is, \( g = \lceil S \rceil \)), then

\[
g = S + 2P/s(s + 1).
\]

All blocks have lengths \( s + 1 \) and \( s + 2 \), with the exception of a small number that can be determined from the relation

\[
\lceil S \rceil - S = 2P/s(s + 1),
\]

where \( P \) is given by (8). This relation guarantees that the number of rogue blocks is very small, and that their lengths are close to those of blocks of lengths \( s + 1 \) and \( s + 2 \). \( \square \)

References


[3] R.G. Stanton and J.L. Allston, A census of values for \( g^{(1)}(1, 2; v) \), Ars Combinatoria 21 (1985) 203–216.

