EXTREMAL PROBLEMS FOR PAIRWISE BALANCED DESIGNS

Paul Erdös
Hungarian Academy of Sciences
H-1053 Budapest, Hungary

Richard A. Duke, Joel C. Fowler, and Kevin T. Phelps
Georgia Institute of Technology
Atlanta, GA 30332

1. Introduction

In a previous paper [2] the senior author discussed many problems concerning block designs. Some of these have since been settled. Szemerédi and Trotter [5] showed, for example, that the number of pairwise balanced designs which can be obtained geometrically on n points in the plane with the blocks being the lines joining these points is at most $\exp(c\sqrt{n})$. In [1] Colbourn, Phelps, and Rödl verified conjecture (2) of [2] by showing that the number of multisets, or "sequences," of integers which can be realized as the block sizes of a pairwise balanced design of order n is between $\exp(c_1\sqrt{n}\log n)$ and $\exp(c_2\sqrt{n}\log n)$. They also found that the number of sets of (distinct) integers which can be realized as block sizes, when multiplicities are not considered, is between $\exp(c_1\sqrt{n})$ and $\exp(c_2\sqrt{n})$.

Here we consider some of the other problems stated in [2] and related questions for such designs. In particular we show that the number of multisets which can be realized as the point-degrees of a pairwise balanced design is between $\exp(c_1 n)$ and $\exp(c_2 n)$ and that bounds of the same form exist for the number of sets which can be realized as degrees. We further show that if the set of block sizes to be used in the design is specified, then for nearly every such choice of block sizes it is still possible to realize exponentially many sets or multisets as the degrees. In fact, we establish that the maximum number of distinct block sizes which can occur in a design is $2\sqrt{n} - 2$ and show that even if we require that our design have this number of distinct block sizes, there still exist exponentially many choices for its degrees.

As in (17) of [2] we let $F(n;k)$ denote the maximum over

Congressus Numerantium, **48 (1985)**, pp.55-66

all designs on n points of the sum of the sizes of the k largest blocks. Here we are able to show that $F(n;k) = n + \binom{k}{2}$ whenever $\binom{k}{2} \le n$. Using a result of Wilson given in [2] we also prove that $F(n;t\sqrt{n}) \le (t+c \log t)n$ for $t > 1$. We conjecture that $F(n;t\sqrt{n}) < (t+c)n$. Another bound, useful for k near n and above, is also obtained.

2. Extremal Results

A pairwise balanced design of order n is an n-set S together with a family B of subsets $A_i \subseteq S$, $|A_i| \ge 2$, $1 \le i \le b$, $b \ge 2$, with the property that each (unordered) pair of elements of S is contained in exactly one member of S. Hereafter, we shall call such a structure simply a design. We refer to the elements of B as the blocks of the design. By the degree of a point $x \in S$ we mean the number of blocks in B which contain x.

The degree sequence of a design (S,B) of order n is the multiset of n integers which are the degrees of the points in S. The question of determining the number of distinct multisets which are the degree sequences of designs of order n was raised in [2]. We will denote this number by G(n). The degree set of a design (S,B) is simply the set of integers which are degrees of points in S. We will let g(n) denote the number of distinct degree sets of designs of order n.

Since no pair of points in a design are contained together in more than one block, it follows immediately that the degree of any point of a design of order n is at most n-1, and thus that there exist positive constants c_1 and c_2 such that $G(n) \le \exp(c_1 n)$ and $g(n) \le \exp(c_2 n)$. We begin by showing that lower bounds of this form also exist for both g(n) and G(n).

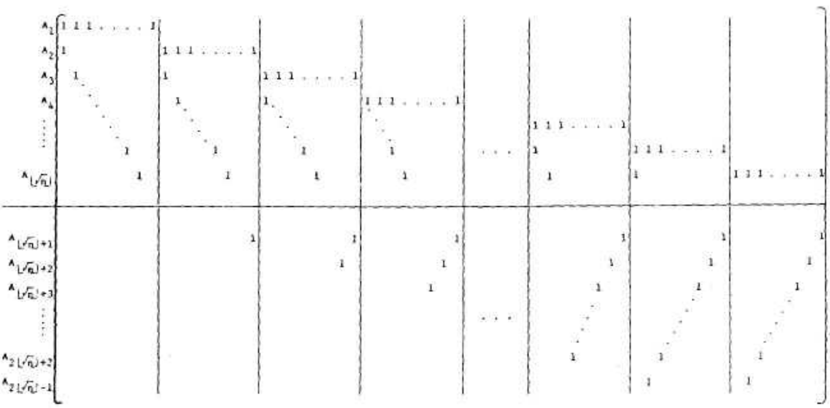
Theorem 1. There exist positive constants c_1, c_2, c_3, and c_4 such that $\exp(c_1 n) \le G(n) \le \exp(c_3 n)$ and $\exp(c_2 n) \le g(n) \le \exp(c_4 n)$.

Proof. As indicated, the upper bounds are immediate. To establish the lower bounds we begin with a Steiner triple system (S,B) of order n. It is well known that such triple systems exist whenever $n \equiv 1$ or $3 \pmod 6$. Each point of S has degree $(n-1)/2$ in this design. Let $x_1, x_2, \ldots, x_m$, $m = \varepsilon n$, $\varepsilon < 1/4$, be points of S. For each i, $1 \leq i \leq m$, select k_i triples, $(n+1)/4 \leq k_i \leq (n-1)/2$, which contain x_i but do not contain any x_j, $j \neq i$, $1 \leq j \leq m$. We delete all of these triples from B. For each pair of points in one of the deleted triples we add that pair as a 2-element block to B. Each point x_i, $1 \leq i \leq m$, has degree $k_i + (n-1)/2 \geq 3n/4$ in the resulting design. Since every point of S $\{x_1, \ldots, x_m\}$ was contained in exactly one of the original triples with any given x_i, the degree of such a point is now at most $\varepsilon n + (n-1)/2 < 3n/4$. The values of the εn k_i's may be selected independently from the interval $[(n+1)/4, (n-1)/2]$, which yields the lower bounds.

Each of the designs constructed in the proof of this theorem contains a positive fraction of the original $n(n-1)/6$ triples and less than $n^2/4$ blocks of size 2. This leads to the following result.

Corollary 2. For sufficiently large n there exists a sequence of block sizes, $|A_1| \geq |A_2| \geq \ldots \geq |A_b|$, such that the number of degree sequences for designs of order n having this multiset of block sizes is at least $\exp(cn)$.

We can modify Theorem 1's construction to obtain exponentially many degree sets for designs in which block sizes are only allowed from some fixed set of at least two positive integers. We illustrate with designs allowing only blocks of two sizes, k_1 and k_2, $(k_1 < k_2)$. The construction is easily generalized to collections of three or more allowable block sizes. First, let v be such that: 1) a design F_1, exists on v points with all blocks of size k_1, and 2) a design, F_2, exists on v points with all blocks of size k_2. Such a v exists by well known theorems of Wilson [6]. Let n be any integer such that there exists a design on n points with all blocks of size v (again, Wilson's results guarantee the existence of such a design for all sufficiently large n which satisfy the elementary necessary conditions on balanced incomplete block designs). We now choose a collection of εn points $x_1, x_2, \ldots, x_{\varepsilon n}$. For any set of positive points. These blocks, together with all two element blocks necessary to cover the remaining paris of points, give a design with the maximum number of different block sizes.



There are $\lfloor\sqrt{n}\rfloor$ column groups, each representing $\lfloor\sqrt{n}\rfloor$ points. Note that A_1 through $A_{\lfloor\sqrt{n}\rfloor}$ give all sizes from $\lfloor\sqrt{n}\rfloor$ to $2\lfloor\sqrt{n}\rfloor-1$ while $A_{\lfloor\sqrt{n}\rfloor+1}$ through $A_{2\lfloor\sqrt{n}\rfloor-1}$ cover sizes 2 through $\lfloor\sqrt{n}\rfloor-1$. Since all uncovered pairs of points will be handled by two element blocks we need only verify that the above blocks cover every pair of points at most once. If two points are chosen from the same $\lfloor\sqrt{n}\rfloor$ group then they occur together in the "major" block defining that group, and no other since the 1's below that block occur at most one to a row. If two points are chosen from different $\lfloor\sqrt{n}\rfloor$ blocks then the structure of the matrix shows that the "left-most" point is in a block (the major block defining that point's $\lfloor\sqrt{n}\rfloor$ group) which the other is not. But every point is in at most two of the blocks of the matrix. Hence this pair is covered at most once.

We note that in the above argument each point was on at most two of the blocks A_i, $1 \le i \le 2\lfloor\sqrt{n}\rfloor-2$. It follows that this construction could be carried out in the plane and hence that this result holds for geometric designs.

integers $b_1, b_2, \ldots, b_{\varepsilon n}$ we replace b_i of the blocks on x_i by the F_1 design. All of the remaining size v blocks are replaced by the F_2 design. By elementary counting it can then be shown that for $b_i \leq (n-1)/(v-1) - \varepsilon n$ the degrees of x_i, $i = 1,2,\ldots,\varepsilon n$ are below the degrees of all remaining points. Hence for ε sufficiently small (in comparison to $\frac{1}{v-1}$) we can obtain any set of εn degrees from a set of size cn for some constant c. This yields exponentially many degree sets.

Our next theorem gives an (exact) bound on the number of different block sizes possible in a design of order n. We will see that even when this wide range of blocks sizes is required exponentially many degree sequences are realizable.

Theorem 3. In a design on n points there can occur no more than $2\lfloor\sqrt{n}\rfloor - 2$ different block sizes. Moreover, for all n a design exists with this number of different block sizes. ($\lfloor x \rfloor$ denotes the integer part of x).

Proof. We show that in a design there can occur no more than $\lfloor\sqrt{n}\rfloor$ blocks of size $\geq \sqrt{n}$. This, together with the fact that no more than $\lfloor\sqrt{n}\rfloor - 2$ sizes are possible from 2 to $\sqrt{n} - 1$, will establish the first part of the theorem.

Let $A_1, A_2, \ldots, A_k$ be blocks of distinct sizes, all $\geq \sqrt{n}$. Then by inclusion/exclusion we have

$$n \geq |A_1 \cup A_2 \cup \ldots \cup A_k| \geq \sum_{i=1}^{k} |A_i| - \sum_{i<j} |A_i \cap A_j|$$

$$\geq \sum_{i=0}^{k-1} (\sqrt{n} + i) - \binom{k}{2}$$

$$= \sqrt{n}\,(k) + \binom{k}{2} - \binom{k}{2}$$

Hence $\lfloor\sqrt{n}\rfloor \geq k$ and the result follows. We note that the above argument shows that a design with $2\lfloor\sqrt{n}\rfloor - 2$ different block sizes must have as those distinct sizes $2,3,4,\ldots, 2\lfloor\sqrt{n}\rfloor - 1$.

For the second part of the theorem we present in Fig. 1 the rows and columns of an incidence matrix which correspond to blocks of sizes $2,3,\ldots,2\lfloor\sqrt{n}\rfloor - 1$ in a design on n

Next we show that exponentially many degree sequences can be realized by designs having the maximum number of block sizes given in Theorem 3.

Theorem 4. Let $h(n)$ the number of degree sequences possible for designs of order n which have at least one block of each size r in the interval $2 \le r \le 2\sqrt{n}-1$. Then $h(n) \ge \exp(cn)$.

Proof. Our approach is based on a combination of the constructions used for Theorems 1 and 3. As for Theorem 1 we begin with a Steiner triple system (S,B) of order n. We add to B blocks A_i, $i = 1,2,\ldots, 2\sqrt{n}-1$, with $|A_i| = i$, constructed as in the proof of Theorem 3. We delete each triple of B which meets any A_i in more than one point and replace each such triple by three pairs. Let B^* be the resulting collection of blocks consisting of the A_i, the remaining triples of B, and these new pairs. Note that the degree of any point in the new design (S,B^*) is between $(n+1)/2$ and $n-1$. Now, as in the proof of Theorem 1, we replace some of the remaining triples by pairs.

Let the points of S be $x_1,x_2,\ldots,x_n$ and let t_i denote the number of triples of $B\backslash B^*$ containing x_i. That is, t_i is the number of triples at x_i which were lost when we added the A_i. The degree of x_i in (S,B^*) is then $t_i + (n+1)/2$ or $t_i + (n+3)/2$. Since each triple of $B\backslash B^*$ meets some A_i in at least two points we have

$$\sum_{i=1}^{n} t_i \le \sum_{j=2}^{2\sqrt{n}-1} \binom{j}{2} \le \frac{4}{3} n^{3/2} .$$

It follows that there exist positive constants c_1 and c_2, $c_1 > c_2$, such that the number of x_i for which $t_i \ge n/5$ is less than $c_1\sqrt{n}$, and the number for which $t_i \ge n/10$ is less than $c_2\sqrt{n}$. We also have that the number of x_i with $t_i < 2\sqrt{n}$ is at least $n/3$. Suppose then that $t_i < 2\sqrt{n}$ for each i, $1 \le i \le m$, $m = \varepsilon n$, $\varepsilon < 1/10$. For each of these x_i select k_i triples in B^* which contain x_i, contain no x_j, $j \ne i$, $1 \le j \le m$, and contain no x_j, $j > m$, for which $t_j \ge n/10$. Note that x_i is on at least $(2n/5) - c_2\sqrt{n}$ such triples. For each i, $1 \le i \le m$, we delete these triples from B^* and replace them with pairs. If each k_i is chosen to be at least $n/4$, then the resulting degree of x_i, $1 \le i \le m$, is at least $3n/4$. Furthermore, the only other points affected are among those with $t_i < n/10$ and so now have degrees less than $3n/4$. Finally we may choose the value of each k_i so that the new degree of x_i is not equal to that of any point for which $t_i \ge n/5$. (There are at most $c_1\sqrt{n}$ such points.) It then follows that each such selection of values for the m k_i's yields a distinct degree sequence and the result follows.

We now turn our attention to the following problem mentioned in [2]. Let

$$F(n;k) = \max\left(\sum_{i=1}^{k} |A_i|\right),$$

where this maximum is taken over all designs possible on n points. Thus $F(n;k)$ is simply the maximum value over all designs on n points of the sum of the k largest block sizes. In [2] it is noted that

$$F(n;k) < \max(c_1 nk^{1/2},\ c_2 n^{1/2} k)$$

for some constants c_1 and c_2.

Here we present tighter bounds on $F(n;k)$.

Theorem 5. $F(n;k) = n + \binom{k}{2}$ for $\binom{k}{2} \le n$.

Proof. We first show $F(n;k) \le n + \binom{k}{2}$ and then construct a design on n points which achieves this bound.

For any set of k blocks in a design we have, arguing as in Theorem 2,

$$n \ge |A_1 \cup A_2 \cup \dots \cup A_k| \ge \sum_{i=1}^{k} |A_i| - \sum_{1\le i<j\le k} |A_i \cap A_j|$$

$$\ge \sum_{i=1}^{k} |A_i| - \binom{k}{2}$$

which immediately gives the bound $F(n;k) \le n + \binom{k}{2}$.

To construct a design achieving this bound we first note that it is always possible to place k blocks of size k-1 onto a set of $\binom{k}{2}$ points such that no pair of points is covered more than once. The incidence matrix for such a configuration is

$$
\begin{array}{c}
A_1\\ A_2\\ A_3\\ A_4\\ \vdots\\ A_{k-1}\\ A_k
\end{array}
\left[\begin{array}{c|c|c|c|c|c}
1\;1\;1\;1\;\cdots\;1 & & & & & \\
1 & 1\;1\;1\;\cdots\;1 & & & & \\
\;\;1 & 1 & 1\;1\;1\;\cdots\;1 & & & \\
\;\;\;\;1 & \;\;1 & 1 & 1\;1\;\cdots\;1 & & \\
\ddots & \ddots & \ddots & \ddots & \cdots\cdots & \\
1 & 1 & 1 & 1 & & 1 \\
\;\;1 & \;\;1 & \;\;1 & \;\;1 & & 1
\end{array}\right]
$$

The first group of points is of size k-1, the second of size k-2, the third of size k-3, etc. Hence the total number of points used is $1 + 2 + \ldots + (k-2) + (k-1) = \binom{k}{2}$. For $n \geq \binom{k}{2}$ we can place the above configuration on $\binom{k}{2}$ of the points and adjoin the remaining $n - \binom{k}{2}$ points to A_1. Add size two blocks to cover any uncovered pairs of points that remain. We now have a design with $|A_1| = (k-1) + n - \binom{k}{2}$ and $|A_i| = k-1$ for $2 \leq i \leq k$. Hence

$$\sum_{i=1}^{k} |A_i| = n + \binom{k}{2}.$$

We note that the above bound and construction are applicable to sets of points in the plane. Thus the above theorem also holds when the designs under consideration are to come from points and lines in the plane.

For larger k in relation to n our results are less exact.

Theorem 6. For $\binom{k}{2} > n$, $F(n;k) \leq k\sqrt{n} + cn \log(k/\sqrt{n})$, for some constant c.

Proof. We rely upon the following bound. If $\mathcal{F}$ is a family of subsets of an n-set, cardinalities from r to s, $\sqrt{n} < r < s$, no pair of which meet in more than one point, then

$$|\mathcal{F}| \leq \frac{n(s-1)}{r^2 - 1} .$$

This can be proved in the same way that bound (6), due to Wilson, in [2] pg. 6, is obtained.

Using the above with $r = (1 + \frac{1}{2^{i+1}})\sqrt{n}$, $s = (1 + \frac{1}{2^i})\sqrt{n}$, and some algebra we obtain that the number of blocks in a design with sizes in this range is no more than $2^{i+1}\sqrt{n}$. Hence the maximum contribution to $F(n;k)$ from blocks with sizes in this interval is

$$\sqrt{n}(1 + \frac{1}{2^i})(\sqrt{n}\ 2^{i+1}) = 2n(2^i + 1),$$

occurring when there are $\sqrt{n}\ 2^{i+1}$ of these blocks. Hence if k_1 is the number of blocks of size $\le 2\sqrt{n}$ and

$$k_1 \sim \sum_{i=0}^{e} \sqrt{n}\ 2^{i+1}$$

then the sum over these k_1 block sizes is bounded by $k_1\sqrt{n}$ + $cn \log(k_1/\sqrt{n})$, which is of the form of the bound of the theorem.

For block sizes larger than $2\sqrt{n}$ we note that no more than $\sqrt{n}$ blocks exist of this size since otherwise

$$n \geq |A_1 \cup \ldots \cup A_{\sqrt{n}}| \geq \sum_{1}^{\sqrt{n}} |A_i| - \binom{\sqrt{n}}{2}$$

$$\geq \sqrt{n}(2\sqrt{n}) - \frac{n - \sqrt{n}}{2},$$

a contradiction. We then apply the bound of Theorem 5 to obtain a maximum contribution of $n + \binom{\sqrt{n}}{2}$ to $F(n;k)$ from blocks of size $\ge 2\sqrt{n}$. This term is easily absorbed into the bound of the theorem.

Our last bound on $F(n;k)$ improves Theorem 6 for k near n and above.

Theorem 7. $F(n;k) \leq (k + \sqrt{k^2 + 4kn(n-1)})/2$.

Proof. Let $A_1, A_2, \ldots, A_k$ be the k largest blocks of a design on n points. Then

$$\sum_{i=1}^{k} |A_i|(|A_i|-1) \leq n(n-1).$$

Hence

$$(\sum_{i=1}^{k} |A_i|)^2 - n(n-1) \leq \sum_{i=1}^{k} |A_i|.$$

By the Cauchy-Schwartz inequality

$$k(\sum_{i=1}^{k} |A_i|)^2 \leq \sum_{i=1}^{k} |A_i|^2 .$$

Thus

$$\left(\sum_{i=1}^{k} |A_i|\right)^2 - k\left(\sum_{i=1}^{k} |A_i|\right) - kn(n-1) \le 0.$$

This quadratic inequality then implies the bound of the theorem.

We note that equality in the above bound implies first that $A_1,\ldots,A_k$ cover all pairs of points and second that all blocks are of the same size. That is, equality holds if and only if we have all blocks of a balanced incomplete block design.

3. Further Problems

Many additional problems on block designs are discussed in [2]. Others can be found in [3]. Here we repeat a few of these questions and pose some new problems which we find particularly interesting.

In [2] it was remarked that "it is perhaps not reasonable to expect to obtain a necessary and sufficient condition for a sequence $x_1,\ldots,x_n$ that there should be a block design" with these block sizes. It was later shown in [1] that the problem of deciding whether a given multiset can be realized in this way is in fact NP-complete. Is there any hope of characterizing those multisets which are the degrees of a design? What about degree sets?

In [4] Larson and Erdös asked whether there exists an absolute constant C such that for each n there exists a pairwise balanced design $\{A_i\}$ of order n such that $|A_i| > n^{1/2} - C$ for all i. Their guess was that no such constant could exist.

For each design consider the maximum number of blocks of any single size. Let f(n) be the minimum of these values over all designs of order n. First we would like to determine whether $f(n) \le cn^{1/2}$. That is, does there exist a design of order n such that for each r the number of blocks of size r is less than $cn^{1/2}$? If true, then apart from the value of c, this result is best possible. Perhaps there exists a design in which for each r the number of blocks of size r is less than $cr^{1/2}$. If such a design does exist it would probably be quite different from those constructed here.

Many open questions remain concerning sums of block sizes. We conjecture that Theorem 6 can be improved to $F(n;k) \le k\sqrt{n} + cn$, for some constant c. Further questions concerning $F(n;k)$ might be formulated as follows. Call a design on n points strong if it has b blocks $(A_1, A_2, \ldots, A_b)$ and for any other design on n points with at least b blocks, largest blocks $B_1, B_2, \ldots, B_b$, we have

$$\sum_{i=1}^{b} |B_i| \le \sum_{i=1}^{b} |A_i|.$$

Theorem 7 and the remarks following its proof show that every balanced incomplete block design is strong. Can anything further be said regarding strong designs?

As we have noted the results of Theorems 3 and 5 are also applicable for those designs which can be obtained from n points in the plane where the blocks are the lines joining these points. Clearly all of our questions about degrees and block sizes could be formulated for such geometric designs, and it would be interesting to determine how the values obtained in these cases compare with those which result when all designs are considered.

By an r-design we understand a set S and a collection of subsets of S such that each r-tuple of S is contained in one and only one of these subsets. Here we have been concerned only with the case $r = 2$. Similar questions could, of course, be asked for r-designs with $r \ge 3$. As far as we know few of these problems have been considered.

4. References

1. C. Colburn, K. Phelps, and V. Rödl, Block sizes in pairwise balanced designs, Canad. Math. Bull. 27 (3) (1984), 375-380.

2. P. Erdös, Problems and results on block designs and set systems. In Proc. of the 13th Southeastern Conf. on Combinatorics, Graph Theory and Computing, Congressus Num. XXXV, Utilitas Math., Winnipeg, 1982, pp. 3-16.

3. P. Erdös, On some problems in graph theory, combinatorial analysis and combinatorial number theory. In Graph Theory and Combinatorics, Proc. of the Cambridge Conference in Honor of Paul Erdös (B. Bollobás, Ed.), Academic Press, London, 1984, pp. 1-17.

4. P. Erdös and J. Larson, On pairwise balanced block designs with the sizes of blocks as uniform as possible, Ann. Discrete Math. 15 (1983), 129-134.

5. E. Szemerédi and T. Trotter, Extremal problems in discrete geoemtry, Combinatorica 3 (3-4) (1983), 381-392.

6. R. M. Wilson, An existence theory for pairwise balanced designs, I and II, Journal of Comb. Theory (A) 13 (1972), 221-273.