Annals of Discrete Mathematics 15 (1982) 129–134

ON PAIRWISE BALANCED BLOCK DESIGNS WITH THE SIZES OF BLOCKS AS UNIFORM AS POSSIBLE

P. ERDÖS and J. LARSON

Dedicated to Prof. N.S. Mendelsohn on his 65th birthday

Let $|\phi| = n$, $A_i \subset \phi$, $1 \leqslant i \leqslant T_n$ is a partially balanced block design, $|A_1| \leqslant \cdots \leqslant |A_{T_n}|$. The authors prove that there is such a design for which $|A_i| = n^{1/2} + O(n^{1/2-c})$ for some $c > 0$.

If certain plausible assumptions on the difference of consecutive primes are made, then the above inequality can be improved to $|A_i| = n^{1/2} + O((\log n)^2)$. It is true that there is a design with $|A_1| > n^{1/2} - c$? This challenging problem is left open.

Let $|S| = n$, $A_i \subset S$, $1 \leqslant i \leqslant m$, $2 \leqslant |A_i| < n$. Assume that every pair (x, y) of elements of S is contained in one and only one A_i. A well-known theorem of de Bruijn and Erdös [1] states that then $m \geqslant n$ where the equality holds if and only if $|A_n| = n-1$, $|A_i| = 2$, $1 \leqslant i < n$, or if the A_i are the lines of a finite geometry. Such a geometry can only exist if $n = u^2 + u + 1$, $|A_i| = u + 1$. Its existence has been established only if u is a prime or a power of a prime. It is one of the outstanding problems of combinatorial mathematics to prove (or disprove) that such a system can only exist if $u = P^\alpha$. Here we want to construct a pairwise balanced design which in some sense is as close to a finite geometry as possible. In fact we prove the following theorem.

Theorem 1. *There is an absolute constant c so that for every sufficiently large n there is a pairwise balanced design for $|S| = n$ with the blocks $A_i \subset S$ satisfying*

$$|A_i| = n^{1/2} + O(n^{1/2-c}), \quad 1 \leqslant i \leqslant m. \tag{1}$$

We will give two proofs for Theorem 1, the first one is constructive and the second one probabilistic which in some sense is more illuminating. But before we prove Theorem 1 we make a few remarks and state some open problems. First of all observe that (1) implies

$$n \leqslant m \leqslant n + O(n^{1-c}). \tag{2}$$

To show (2), observe that, since every pair of elements of S must be contained in one and only one A_i, we have

$$\sum_{i=1}^{m} \binom{|A_i|}{2} = \binom{n}{2},$$

and thus the upper bound of (2) immediately follows from (1). The lower bound follows from the theorem of de Bruijn and Erdös.

The following problem is interesting but seems difficult: Does there exist a pairwise balanced design satisfying

$$|A_i| = n^{1/2} + O(1). \tag{3}$$

If (3) holds, then as in (2) we would have $n \leqslant m < n + c_1 n^{1/2}$.

At the moment we do not see how to decide (3), but we will show that if we make certain plausible (but hopeless) assumptions on the difference of consecutive primes, then we obtain the slightly weaker

$$|A_i| = n^{1/2} + O((\log n)^2). \tag{4}$$

Constructive proof of Theorem 1. Let p_k be the smallest prime for which $p_k^2 + p_k + 1 \geqslant n$. A well-known theorem of Iwaniec and Heath-Brown [3] states that, for $k > k_0(\varepsilon)$,

$$p_{k+1} - p_k < p_k^{11/20+\varepsilon}. \tag{5}$$

Eq. (5) implies that if p_k is the smallest prime for which $p_k^2 + p_k + 1 \geqslant n$, then

$$n \leqslant p_k^2 + p_k + 1 < n + n^{31/40+\varepsilon}. \tag{6}$$

Let now $|S_1| = p_k^2 + p_k + 1$ and consider a finite projective Desarguesian plane on S_1. Let $L_1, \ldots, L_{p_k^2+p_k+1}$, $|L_i| = p_k + 1$ be the lines of S_1 and let C be a conic of our geometry. Let x be a point not on C and $L_1, \ldots, L_{p_k+1}$ the lines through x, let further $L_1, \ldots, L_{(p_k-1)/2}$ be the lines which do not meet C. Put

$$p_k^2 + p_k + 1 - n = rp_k + 1 + s, \quad 0 \leqslant s < p_k,$$

and by (6), $0 \leqslant r < p_k^{11/20+\varepsilon}$.

Omit now from S_1 the lines $L_1, \ldots, L_r$ and all the $r\,p_k + 1$ points on it and also omit s points of our conic C (a conic has $p_k + 1$ points). Thus we are left with a set S of n elements. The lines $L_1, \ldots, L_r$ disappeared, if $r < j \leqslant p_k^2 + p_k + 1$, then we now determine how many points we omitted from L_j. If

$r < j \leq p_k + 1$, i.e., if $x \in L_j$, then we omitted one, two or three points of L_j. To see this observe that x has been omitted and if L_j does not meet C we only omitted one of its points. If it meets C, then perhaps one or two more of its points have been omitted. If $x \notin L_j$ (or $p_k + 2 \leq j \leq p_k^2 + p_k + 1$), then we certainly omitted at least r points from L_j (since it meets each of the lines L_i, $1 \leq i \leq r$ in one point) and perhaps we omitted one or two of the points $L_j \cap C$. Let us now denote by A_{j-r} what remains from L_j after omitting our points, $(r < j \leq p_k^2 + p_k + 1)$. The sets $A_1, \ldots A_{p_k^2+p_k+1-r}$ clearly give a pairwise balanced design of the set S, $|S| = n$ and there are at most six possible values of $|A_j|$, namely

$$p_k - 1, \quad p_k - 2, \quad p_k - 3, \quad p_k - r,$$
$$p_k - r - 1, \quad p_k - r - 2, \quad r < p_k^{11/20+\varepsilon}.$$

This completes the proof of Theorem 1.

Probabilistic proof of Theorem 1. We shall show that if we omit from S_1 in all possible ways

$$T_n = p_k^2 + p_k + 1 - n$$

elements, we almost surely are left with a set S, which will satisfy (1). We can omit T_n elements from S_1 in

$$\binom{p_k^2 + p_k + 1}{T_n}$$

ways. To complete our proof it will suffice to show that for all but

$$\mathrm{o}\left(\binom{p_k^2 + p_k + 1}{T_n}\right)$$

of these omissions, we omitted from each L_i, $1 \leq i \leq p_k^2 + p_k + 1$,

$$T_n/p_k + \mathrm{o}((T_n/p_k)^{1/2}(\log n)^2) \tag{7}$$

elements. Eq. (7) easily follows by standard methods of elementary theory of probability and we only outline the proof. Put

$$\frac{T_n}{p_k} + \varepsilon\left(\frac{T_n}{p_k}\right)^{1/2}(\log n)^2 = u, \qquad \frac{T_n}{p_k} - \varepsilon\left(\frac{T_n}{p_k}\right)^{1/2}(\log n)^2 = v.$$

Then the number of ways we can omit from $p_k^2+p_k+1$ elements T_n of them so that there should be at least one line L_i, $1\le i\le p_k^2+p_k+1$ from which we omitted more than u or fewer than v elements is clearly less than

$$(p_k^2+p_k+1)\left(\sum_{s=1}^{p_k+1-u}\binom{p_k+1}{u+s}\binom{p_k^2}{T_n-u-s}+\sum_{t=1}^{v}\binom{p_k+1}{v-t}\binom{p_k^2}{T_n-v+t}\right). \tag{8}$$

A simple computation which we suppress gives that the expression in (8) is

$$\mathrm{o}\left(\binom{p_k^2+p_k+1}{T_n}\right)$$

which again completes the proof of Theorem 1.

At the moment we do not see how to prove (or disprove) (3). The constructive proof of Theorem 1 gave a pairwise balanced design with only 6 different sizes of the blocks. It would be of some interest to show that 6 can be decreased to 3 and perhaps even to 2.

Now we deduce (4) from conjectures on $p_{k+1}-p_k$. The Riemann hypothesis would imply $p_{k+1}-p_k<p_k^{1/2+\varepsilon}$ and nearly 100 years ago Piltz conjectured $p_{k+1}-p_k=\mathrm{o}(p_k^{\varepsilon})$. Finally 50 years ago Cramer [2] conjectured that

$$\limsup(p_{k+1}-p_k)/(\log k)^2=1\,. \tag{9}$$

Eq. (9) seems to be unattackable by the techniques at our disposal. We now deduce (4) from (9). First we prove the following lemma.

Lemma 2. *In every finite geometry of p^2+p+1 points there always is a set of lines $L_1,\dots L_r$, $r\ge p^{1/5}$ so that no three of the L_i are concurrent and no three of the $\binom{r}{2}$ points $L_i\cap L_j$, $1\le i<j\le r$ are on a line.*

Proof. The proof of Lemma 2 is simple. Let $L_1,\dots,L_r$ be a maximal system of lines satisfying the conditions of Lemma 2. In other words if L_u is any of the other $p^2+p+1-r$ lines of our geometry L_u either goes through one of the $\binom{r}{2}$ points $L_i\cap L_j$, $1\le i<j\le r$ or for some k, i_1, j_1, i_2, j_2, $1\le k\le r$, $1\le i_1<j_1\le r$, $1\le i_2<j_2\le r$ the points $L_u\cap L_k$, $L_{i_1}\cap L_{j_1}$, $L_{i_2}\cap L_{j_2}$ are on a line. The first condition eliminates at most $\binom{r}{2}(p+1)$ lines and the second condition

$$(p+1)r\binom{\binom{r}{2}}{2}$$

lines. Thus by our maximality condition we must have

$$p^2+p+1-r \leqslant (p+1)\left(\binom{r}{2}+r\binom{\binom{r}{2}}{2}\right)$$

or $r > p^{1/5}$ which proves Lemma 2.

Let C be a conic of our geometry. Observe that Lemma 2 remains true if we further insist that none of our lines $L_1, \ldots, L_r$ intersect. The proof of this follows immediately from the fact that there are

$$p^2+p+1-(p+1)-\binom{p+1}{2}=\binom{p}{2}$$

lines not intersecting C.

Now we are ready to deduce (4) from (9). Let as in the proof of Theorem 1 p_k be the smallest prime for which $p_k^2+p_k+1 \geqslant n$. Eq. (9) implies that for $n > n_0$

$$n \leqslant p_k^2+p_k+1 < n+3(\log n)^2 . \tag{10}$$

Let r be the largest integer for which

$$p_k^2+p_k+1-n > p_k+1+p_k+(p_k-1)+\cdots+p_k-r$$

and put

$$p_k^2+p_k+1-n=2p_k+1+\sum_{i=1}^{r}(p_k-i)+s, \quad 0<s<p_k-r-1$$

and by (10) $r \leqslant 3(\log n)^2$. Let now $|S_1| = p_k^2+p_k+1$ be a finite geometry and $L_1, \ldots, L_{r+2}$ are $r+2$ lines which satisfy Lemma 2 and do not meet the conic C. Omit the lines $L_1, \ldots, L_{r+2}$ and all the points on them and also s points of the conic C. Then we are left with a pairwise balanced design on S, $|S| = n$ with

$$p_k^2+p_k-r-1=n+\mathrm{O}(n^{1/2}(\log n)^2)$$

blocks A_i, $1 \leqslant i \leqslant p_k^2+p_k-r-1$. By Lemma 2 a line L_j, $j \neq 1, 2, \ldots, r+2$ meets $\bigcup_{i=1}^{r+2} L_i$ in at most $r+2$ and at least r points, further L_j can meet C in 0, 1 or 2 points. Thus the possible values of $|A_i|$ are

$$p_k+1-r, \qquad p_k-r, \qquad p_k-r-1, \qquad p_k-r-2, \qquad p_k-r-3,$$

which by (10) proves (4). Our method is quite inadequate for the proof of (3) and if (3) is true a new idea will probably be required.

The following problem is perhaps of some interest. Consider a finite geometry of $n = u^2 + u + 1$ points. Let $x_1, \ldots, x_k$ be a maximal set of points no three of which are on a line. In other words the lines joining x_i and x_j, $1 \leq i < j \leq k$ contain all the points of our geometry. Determine or estimate the smallest possible value of k. Clearly $k > n^{1/4}$. Is $k = o(n^{1/2})$ possible ? Can the exponent $\frac{1}{5}$ in Lemma 2 be improved ?

University of Florida
Gainesville
Florida, USA

References

[1] N.G. de Bruijn and P. Erdös, On a combinatorial problem, Nederl. Akad. Wet. 51 (1948) 1277–1279.

[2] H. Cramer, On the order of magnitude of the difference between consecutive prime number, Acta Arith. 2 (1936) 23–46.

[3] D.R. Heath-Brown and H. Iwaniec, On the difference between consecutive primes, Bull. Amer. Math. Soc. (N.S.) 1 (1979) 758–760.