INTERSECTION PROPERTIES OF FAMILIES
CONTAINING SETS OF NEARLY THE SAME SIZE

P. Erdös, R. Silverman and A. Stein

Abstract

A family F of sets has property $B(s)$ if there exists a set S whose intersection with each set in F is non-empty but contains fewer than s elements.

P. Erdös has asked whether there exists an absolute constant c such that every projective plane has property $B(c)$.

In this paper, the authors, as a partial answer to this question, obtain the result that for n sufficiently large, every projective plane of order n has property $B(c \log n)$. The result is a corollary of a theorem applicable to somewhat more general families of finite sets.

A family F of sets has property B if there is a set S whose intersection with each set in the family is a proper subset of that set. Many algebraic and combinatorial problems may be restated in terms of property B.

P. Erdös [1] proposed property $B(s)$ as a stronger form of property B. A family F has property $B(s)$ if there is a set S whose intersection with each set in the family is a proper subset of that set containing fewer than s elements. Property $B(s)$ has been studied by, among others, Silverman, Levinson, Stein, Abbott and Erdos.

Property B is an important combinatorial property, and the related literature includes, for example, a recent paper by D. Kleitman on Sperner families [4].

In this paper we consider 2 questions:

1) P. Erdös has asked whether there exists an absolute constant c such that every projective plane has property $B(c)$.

2) Further, can any analogous result be obtained for families of sets more general than projective planes?

ARS COMBINATORIA, Vol. 15 (1983), pp. 247-259

Using probabilistic methods, we obtain a partial result for projective planes, and a somewhat more general result for finite sets. The results obtained show that there is some constant c such that, for n large enough, every projective plane of order n has property $B(c \log n)$. This is done in Section I.

We also obtain the result that a projective plane of order n has property $B(n - c\sqrt{n})$. Although this is a weaker result than the one above, the proof is of interest because it is constructive. This is done in Section II.

Section I: Our main result is the following.

THEOREM 1. *Let* $0 < a_1 \le a_2$, $0 < b$. *Suppose* $0 \le \delta \le 1$, $s \ge 1$ *if* $\delta = 0$, $s < 0$ *if* $\delta = 1$. *Then for any fixed* c_1, *there is some* c_2 *such that if* $\mathcal{F}$ *is a family of sets satisfying the following conditions:*

i) $a_1 n \le |F| \le a_2 n$ *for every* $F \in \mathcal{F}$

ii) $|\mathcal{F}| \le n^b$

then there is some set S *such that if* $F \in \mathcal{F}$ *then*

(1) $$c_1 n^\delta \log^s n \le |S \cap F| \le c_2 n^\delta \log^s n.$$

It will further be shown that if $\delta > 0$, $s > 1$ *or* $c_1 > eb(a_2/a_1)$ *and we restrict ourselves to large* n, *then* c_2 *can be chosen arbitrarily close to* c_1, *while otherwise (again for large* n*)* c_2 *can be chosen arbitrarily close to* $\frac{a_2}{a_1} eb$.

The proof requires certain lemmas, used for bounding tails of the multinomial and binomial distributions, which are of independent interest.

The first lemma relates the tail of a binomial distribution to the largest term of the tail.

LEMMA 1. *Let* $0 < p < 1$, $q = 1 - p$.

(2) *Let*
$$t(x_0) = \begin{cases} \sum_{x \ge x_0} \binom{z}{x} p^x q^{z-x} & \text{if } x_0 > zp \\ \sum_{x \le x_0} \binom{z}{x} p^x q^{z-x} & \text{if } x_0 \le zp. \end{cases}$$

Then $t(x) \le x_0 \binom{z}{x_0} p^{x_0} q^{z-x_0}$.

Proof. If $x_0 \le zp$, then the bound is trivial. Thus we shall concentrate on the case where $x_0 > zp$. We use the following well-known identity.

(3)
$$t(x_0) = z\binom{z-1}{x_0-1} \int_0^p t^{x_0-1}(1-t)^{z-x_0} dt.$$

A simple differentiation shows that the integrand attains its maximum at $t = p$. The result follows immediately.

The next lemma relates the size of a binomial coefficient $\binom{M}{N}$ to the fraction N/M.

LEMMA 2. *There exists an absolute constant* β *such that, if* $0 < N < M$, *then*

(4)
$$\binom{M}{N} \le \frac{\beta}{(\frac{N}{M})^N (1 - \frac{N}{M})^{M-N}} .$$

Proof. By Sterling's formula, there is some constant β' such that

(5)
$$\binom{M}{N} \le \beta' \left(\frac{M^{M+\frac{1}{2}}}{e^M}\right)\left(\frac{e^N}{N^{N+\frac{1}{2}}}\right)\left(\frac{e^{M-N}}{(M-N)^{M-N+\frac{1}{2}}}\right) .$$

Thus

(6)
$$\binom{M}{N} \le \sqrt{\frac{M}{M-1}} \beta' \frac{M^M}{N^N (M-N)^{M-N}} ,$$

and (4) follows.

We also need to bound the terms of a multinomial distribution using terms of a binomial distribution.

LEMMA 3. *Let* $\gamma < 1$, $A + B = C$, $x + y = z \le \gamma C$, $p = A/C$, $q = B/C = 1 - p$, $h(x) = \binom{A}{x}\binom{B}{y}/\binom{C}{z}$, $b(x) = \binom{z}{x}p^x q^y$. *Then* $h(x) = O_\gamma(b(x))$.

Proof. Let

$$(7)\qquad Q = \frac{h(x)}{b(x)} = \frac{(1 - \frac{1}{A})\dots(1 - \frac{x-1}{A})(1 - \frac{1}{B})\dots(1 - \frac{y-1}{B})}{(1 - \frac{1}{C})\dots(1 - \frac{z-1}{C})} .$$

Without loss of generality, we may assume $A \le C/2$. Since $B \le C$, $1 - j/B \le 1 - j/C$, so

$$(8)\qquad Q \le \frac{(1 - \frac{1}{A})\dots(1 - \frac{x-1}{A})}{(1 - \frac{y+1}{C})\dots(1 - \frac{y+x-1}{C})} \cdot \frac{1}{y - \frac{y}{C}} .$$

Since

$$1 - \frac{j}{A} \le 1 - \frac{y+j}{C} \quad \text{if} \quad j \ge \frac{y}{C-A} ,$$

and

$$1 - \frac{j}{A} \le 1,$$

$$Q \le \frac{1}{\prod\limits_{0 \le j < \frac{y}{C-A}} (1 - \frac{y+j}{C})} .$$

Now consider

$$(9)\qquad \log Q \le - \sum_{0 \le j < \frac{y}{C-A}} \log (1 - \frac{y+j}{C}).$$

Since

$$y + j < z \le \gamma C, \ \frac{y+j}{C} < \gamma$$

and

$$-\log(1 - \frac{y+j}{C}) \le \frac{3-2\gamma}{2-2\gamma} \cdot \frac{y+j}{C} .$$

Hence

$$\log Q \le \frac{3-2\gamma}{2-2\gamma} \sum_{0 \le j < \frac{y}{C-A}} \frac{y+j}{C}$$

$$\le \frac{3-2\gamma}{2-2\gamma} \left\{ \frac{y}{C} + \frac{y + \frac{y}{C-A}}{C} \right\} \frac{\frac{y}{C-A} + 1}{2}$$

or

$$\log Q \le \frac{3-2\gamma}{2-2\gamma} \cdot \frac{y}{2(C-A)^2 \cdot C} (2(C-A) + 1)(y + C - A)$$

or

$$\log Q \le \frac{3-2\gamma}{2-2\gamma} \cdot \frac{\gamma C}{2(C/2)^2 \cdot C} \cdot 2C \cdot (1+\gamma)C = O_\gamma(1).$$

We now use these first three lemmas to obtain a bound on the tail of a multinomial distribution.

LEMMA 4. *Let* $\alpha > 0$, $a > 1$, $A = \alpha n$, $B \ge n^a$, $C = A + B$, $z = kBn^\delta \log^s n/n$ *(where* s *and* δ *are restricted as in the statement of the theorem),* $x_0 = cn^\delta \log^s n$. *Let* $p, q, h(x), b(x)$ *be defined as in Lemma 3. Let*

$$(10) \qquad T(x_0) = \begin{cases} \sum_{x \ge x_0} h(x) & \text{if } x_0 > zp \\ \sum_{x < x_0} h(x) & \text{if } x_0 \le zp. \end{cases}$$

Then

$$(11) \qquad T(x_0) \ll \exp\{n^\delta \log^s n (c(1 + \log \frac{\alpha k}{C} + o(1)) - \alpha k + o(1)) + \delta \log n + s \log \log n\}.$$

Proof. Since $\frac{z}{c} \le \frac{kn^\delta \log^s n}{n} \to 0$ as $n \to \infty$, we may use Lemma 3 to obtain $T(x_0) \ll t(x_0)$. We now use the bound on $t(x_0)$ from Lemma 1 to obtain

$$(12) \qquad T(x_0) \ll x_0 \binom{z}{x_0} p^{x_0} q^{z-x_0}.$$

Using the bound on the binomial coefficient obtained in Lemma 2 as well as our definition of p and q, we find

$$(13) \qquad T(x_0) \ll x_0 \cdot \frac{1}{\left(\frac{x_0}{z}\right)^{x_0}\left(1-\frac{x_0}{z}\right)^{z-x_0}} \left(\frac{A}{C}\right)^{x_0}\left(\frac{B}{C}\right)^{z-x_0} .$$

Segregating by exponent,

$$(14) \qquad T(x_0) \ll x_0 \left(\frac{z(1-\frac{x_0}{z})A}{x_0 B}\right)^{x_0} \left(\frac{B}{C(1-\frac{x_0}{z})}\right)^{z}$$

and thus

$$(15) \qquad T(x_0) \ll cn^{\delta}\log^{s}n \; \frac{\left\{\frac{\alpha k}{c}\left(1-\frac{cn}{kB}\right)\right\}^{cn^{\delta}\log^{s}n}}{\left\{\left(1+\frac{an}{B}\right)\left(1-\frac{cn}{kB}\right)\right\}^{\frac{kBn^{\delta}\log^{s}n}{n}}} .$$

Since $B \geq n^{a}$ with $a > 1$,

$$(16) \qquad \left(1-\frac{cn}{kB}\right)^{cn^{\delta}\log^{s}n} \sim \exp\left\{-\frac{c^{2}n^{1+\delta}\log^{s}n}{kB}\left(1+o(1)\right)\right\} ,$$

$$(17) \qquad \left(1+\frac{\alpha n}{B}\right)^{\frac{kBn^{\delta}\log^{s}n}{n}} \sim \exp\left\{\alpha kn^{\delta}\log^{s}n\left(1+o(1)\right)\right\}$$

and

$$(18) \qquad \left(1-\frac{cn}{kB}\right)^{\frac{kBn^{\delta}\log^{s}n}{n}} \sim \exp\{-cn\ \log^{s}n(1+o(1))\}.$$

Combining (15)-(18) yields (11).

We now use this estimate to show that the tail can be made smaller than any negative power of n.

LEMMA 5. *(Let everything not specifically defined below be defined as in Lemma 4.) Suppose α is bounded away from both 0 and ∞, b fixed and c_1 arbitrary. Then there exist k, c_2 such that,* *defining* $x_i = c_i n^{\delta} \log^s n$ *and* $T(x_1)$ *and* $T(x_2)$ *as the lower and upper tails of the multinomial distribution as in Lemma 4,*

$$(19) \qquad n^b T(x_i) = o(1), \quad i = 1,2.$$

Proof. We must show that k, c_2 can be chosen so that

$$(20) \qquad b \log n + n^{\delta} \log^s n (c_i (1 + \log \frac{\alpha k}{c_i} + o(1)) - \alpha k + o(1)) + \delta \log n + s \log \log n \to -\infty$$

as $n \to \infty$, $i = 1,2$.

First choose k large enough to make the coefficient of $n^{\delta} \log^s n$ less that $-(b + \delta)$ for $i = 1$. Then choose c_2 so that the same condition holds with $i = 2$.

We are now prepared to prove a preliminary version of the theorem.

LEMMA 6. *Let* $\alpha > 0$. *If the hypotheses of the theorem hold then the conclusion also holds if* $\mathcal{F}$ *also satisfies the additional condition*

$$(21) \quad iii. \qquad |\bigcup_{F \in \mathcal{F}} F| \geq \max(n^a, n^b).$$

Proof. Let $F^* = \bigcup_{F \in \mathcal{F}} F$. Choose k so that $z = k|F^*| n^{\delta} \log^s n / n$ is an integer. Let E_F be the event that a set $S \subset F^*$ of size z satisfies

$$|S \cap F| < c_1 n^{\delta} \log^s n \quad \text{or} \quad |S \cap F| > c_2 n^{\delta} \log^s n.$$

We will prove the lemma by showing that k and c_2 can be chosen so that $P(E) < 1$.

Letting R_1 represent $<$ and R_2 represent $>$,

$$P(|S \cap F| R_i c_i n^{\delta} \log n) = \sum_{x R_i c_i n^{\delta} \log n} h(x).$$

where

$$(22) \qquad h(x) = \frac{\binom{|F|}{x}\binom{|F^*|-|F|}{z-x}}{\binom{|F^*|}{z}} .$$

Thus

$$(23) \qquad P(|S \cap F| R_i c_i n^{\delta} \log n) = T_F(x_i), \quad i = 1,2$$

and hence

$$(24) \qquad P(E_F) \le T_F(x_1) + T_F(x_2).$$

Now let E be the event that a set $S \subset F^*$ of size z satisfies $|S \cap F| < c_1 n^{\delta} \log^s n$ or $|S \cap F| > c_2 n^{\delta} \log^s n$ for at least one set $F \in F$. Then

$$(25) \qquad P(E) \le \sum_{F \in F} P(E_F) \le \sum_{F \in F} [T_F(x_1) + T_F(x_2)]$$

$$\le |F| \max T_F(x_1) + |F| \max T_F(x_2)$$

$$\le 2n^b \max T_F(x_i) = o(1)$$

by Lemma 5. This proves Lemma 6.

We now observe that condition iii. is unnecessary. If $|F|$ is not large enough, we may augment F by including sets disjoint from the original sets. The conclusion will hold for this augmented family and thus must also hold for the original family F as well. This proves the theorem.

Note that our proof has actually shown that almost all subsets of F^* of size $k|F^*|n^{\delta}\log^s n/n$ will be "blocking sets".

We may observe that if $\delta > 0$ or $s > 1$, then, in the proof of Lemma 5, it is only necessary to make $\log\frac{\alpha k}{c_i} < -1$. Thus k can be chosen arbitrarily close to $c_1/e\alpha_2$ and c_2 arbitrarily close to $e\alpha_2 k$. Thus if $\delta > 0$ or $s > 1$ then c_2 can be taken arbitrarily close to c_1.

If $\delta = 0$ and $s = 1$, then it suffices to make

$$(26) \qquad 1 + \log \frac{\alpha k}{c_i} < 0, \ \alpha k > b$$

for $\alpha_1 \leq \alpha \leq \alpha_2$, $i = 1,2$.

This is equivalent to making

$$(27) \qquad b/\alpha_1 < k < \frac{c_1}{e\alpha_2}.$$

This can be done if $c_1 > eb(\frac{\alpha_2}{\alpha_1})$. In this case c_2 can be made arbitrarily close to c_1. If $c_1 \leq eb(\frac{\alpha_2}{\alpha_1})$, then k must be chosen large forcing c_2 to be chosen larger as well in (20). However, in this case we can certainly choose c_2 close to $eb(\alpha_2/\alpha_1)$. We thus see that if

i) $\delta > 0$, $s > 1$ or $c_2 > eb(\alpha_2/\alpha_1)$

and

ii) $c_2 > c_1$,

then, if n is large enough, there is some set S such that $c_1 n^\delta \log^s n \leq |S \cap F| \leq c_2 n^\delta \log^s n$ for every $F \in \mathcal{F}$.

Another way of looking at the above is that, thinking of c_2 as a function of n,

$$(28) \qquad \lim_n \inf c_2 \begin{cases} = c_1 & \text{if } \delta > 0,\ s > 1 \text{ or } c_1 \geq eb(\alpha_2/\alpha_1), \\ \leq eb(\frac{\alpha_2}{\alpha_1}) & \text{otherwise.} \end{cases}$$

Applying the above to the case of projective planes, we immediately have the following corollary.

COROLLARY. *Let $c > 2e$. If n is large enough, then the projective plane of order n has property $B(c \log n)$.*

Section II.

We now demonstrate construction of a "blocking set" and show that a projective plane P of order n has property $B(n - p(n))$, where $p(n)$ is of order $\sqrt{n}$.

We first indicate the method of proof. Consider an arbitrary point x in P, and the lines $\ell_1,\ldots,\ell_{n+1}$ through x. The lines have the properties that: a) $i \neq j \Rightarrow \ell_i \cap \ell_j = \{x\}$; b) $\bigcup_i \ell_i = P$. To pick the points for the "blocking" set S, we: 1) pick $y_1,\ldots,y_k$, y_i on line ℓ_i, in *general* position, i.e., no line in P containing more than 2 of them (we can do this as long as $\binom{k-1}{2} < n$). 2) repeat 1), k lines at a time. No line contains more than 2k' of $\{y_i\}$, where $k' = [\frac{n}{k}] + 1$, and the set intersects every line through x.

Now, consider a line ℓ, not containing x. Let $\ell = \{x^1,\ldots,x^j, x^{j+1},\ldots,x^{n+1}\}$. Every other line of P contains exactly one point of ℓ. We pick the remaining points for S as follows: 3) repeat 2), for $i = 1,\ldots,j$ where $j \leq \frac{n}{2k'+1}$; 4) augment the set obtained from 1), 2) and 3) by $x^{j+1},\ldots,x^{n+1}$.

The aggregate set S obtained from steps 1) through 4) has the required properties of intersecting each line in P in a non-empty set whose cardinality is less than $n + 2 - j$, so P has property $B(n + 2 - j)$. We further note that $j \sim \sqrt{n}$, so P has property $B(n-p(n))$, where $p(n) \sim \sqrt{n}$. This is the desired result.

LEMMA 7. *Let x be a point in P, and $\ell_1,\ldots,\ell_k$ be k distinct lines through x, where k is a positive integer solution of $(k-1)(k-2) < 2n$. Then we can choose points $y_i \in \ell_i$, $i = 1,2,\ldots,k$, such that no line in P contains more than 2 of the y_i.*

Proof. Choose $y_1 \in \ell_1$, $y_2 \in \ell_2$. There is a line in P, $\langle y_1,y_2\rangle$ containing both y_1 and y_2. ℓ_3 intersects that line in one point, so there are points other than x in ℓ_3 not on $\langle y_1,y_2\rangle$. Let $y_3 \in \ell_3 - \langle y_1,y_2\rangle$. Inductively, select $y_i \in \ell_i$, $i = 1,2,\ldots,k-1$, in such a way that no line of P contains more than two of the collection.

That this is possible, can be seen as follows. When k-1 points have been selected, there are exactly $\binom{k-1}{2}$ lines in P containing two of them. ℓ_k intersects each such line in one point. Since ℓ_k contains n+1 points, there is a point on ℓ_k which is not x, and not on any of those $\binom{k-1}{2}$ lines, as long as $n+1 > \binom{k-1}{2} + 1$. But this condition is assured by the hypothesis that $(k-1)(k-2) < 2n$.

LEMMA 8. *As in Lemma 7, let x be a point in P, and $\ell_1,\ldots,\ell_k$ be k distinct lines through x, where k is a positive integer solution of $(k-1)(k-2) < 2n$. Furthermore, let k' be the smallest integer such that $k' \geq \frac{n}{k}$. Then we can choose points $y_i \in \ell_i$, $i = 1,2,\ldots,n$, such that no line in P contains more than $2k'$ of the y_i.*

Proof. Choose $y_1 \in \ell_1, y_2 \in \ell_2,\ldots,y_k \in \ell_k$ as in Lemma 7. Similarly choose $y_{k+1} \in \ell_{k+1}, y_{k+2} \in \ell_{k+2},\ldots,y_{2k} \in \ell_{2k}$ and then continue, in groups of k points, ultimately reaching $y_{(k'-2)k+1} \in \ell_{(k'-2)k+1},\ldots,$ $y_{(k'-1)k} \in \ell_{(k'-1)k}$. Finally, again using Lemma 7, choose $y_{(k'-1)k+1} \in \ell_{(k'-1)k+1},\ldots,y_n \in \ell_n$. We have partitioned $y_1,\ldots,y_n$ into k' subsets such that no line in P contains more than 2 points from any subset. Thus no line in P contains more than 2k' of the y_i.

LEMMA 9. *Select integers k and k' as in Lemma 8. Let ℓ be a line in P and $x^{(1)},x^{(2)},\ldots,x^{(j)}$ be distinct points on ℓ. We can choose a set $S^{(j)}$ of points in P such that*

a) If ℓ' is a line in P, then no more than $2jk'$ elements in $S^{(j)}$ are on ℓ', and
b) If $\ell' \neq \ell$ is a line in P containing one of the points $x^{(i)}$, then $S^{(j)}$ contains at least one point. of ℓ'.

Proof. For each $x^{(i)}$, let $\ell_1^{(i)},\ldots,\ell_n^{(i)}$ be the lines in P, other than ℓ, containing $x^{(i)}$. For each $x^{(i)}$, choose $y_1^{(i)} \in \ell_1^{(i)}$, $\ldots,y_n^{(i)} \in \ell_n^{(i)}$ as in Lemma 8. Let $S^{(j)}$ be the set of $y_m^{(i)}$ so so chosen. Condition b) is clearly satisfied. So is condition a), as we can partition $S^{(j)}$ into j components and no line $\ell' \neq \ell$ contains more than 2k' points from each component. Note that $S^{(j)}$ is also disjoint from ℓ, since it contains no $x^{(i)}$, and further each point in $S^{(j)}$ is chosen from a line other than ℓ which contains some $x^{(i)}$ and thus no other points of ℓ.

We are now ready to prove the main result.

THEOREM 2. *Select k, k' as in Lemma 8, and an integer $j \leq n/(2k'+1)$. Then P has property $B(n+2-j)$.*

Proof. Choose $S^{(j)}$ as in Lemma 9 and let $S' = \{x \in \ell: x$ is not one of the $x^{(i)}\}$. Let $S = S^{(j)} \cup S'$.

Since each line disjoint with S' is not ℓ and contains one of $\{x^{(i)}\}$, and each such line contains one element of $S^{(j)}$, S contains at least one point on each line. Since S' contains $n+1-j$ points of ℓ and $S^{(j)}$ is disjoint with ℓ, S contains exactly $n+1-j$ points of ℓ.

On the other hand, if $\ell' \neq \ell$, then ℓ' contains at most $2jk'$ points of $S^{(j)}$ and one point of S'. Thus ℓ' contains at most $2k' + 1$ points of S.

But $j \leq n/(2k'+1)$, so $2jk'+1 \leq n+1-j$. Thus no line in P contains more than $n+1-j$ points of S. We further note that $k \sim \sqrt{2n}$ and $k' \sim \sqrt{n/2}$, thus $j \sim \sqrt{n/2}$. Q.E.D.

ACKNOWLEDGEMENT: The authors wish to thank the editor for his helpful suggestions and Joel Spencer, who has pointed out that the necessity for using the multinomial distribution in Section I can be obviated by choosing S in the following manner: For each $x \in F^*$, let x belong to S with probability $kn^{\delta}\log^{s}n/n$.

References

[1] H.L. Abbott and A. Liu, *On property B(s)*, Ars Combinatoria, v. 7, (1979), pp. 250-260.

[2] P. Erdös, *On a combinatorial problem*, Nordisk Mat. Tidskr., v. 11, (1963), pp. 5-10.

[3] P. Erdös, Chau Ko and R. Rado, *Intersection theorems for systems of finite sets*, Quart. J. Math., Oxford (2), 12 (1961), 313-320.

[4] D. Kleitman, *The number of Sperner families of subsets of an n element set*, Colloq. Math. Soc. J. Bolyai, 10, 1973, pp. 989-1001.

[5] H. Levinson, and R. Silverman, *Generalizations of property B*, Proc. 2nd Int. Conf. Comb. Math., NYAS, NY, 1978.

[6] R. Silverman and A. Stein, *Minimal families lacking property B(s) II*, Proc. 11th S.E. Conf. on Comb., Graph Theory, and Computing, Boca Raton (1980).

[7] J. Spencer, *Sequences with small discrepancy relative to n events*, Compositio Mathematica (to appear).

Nemetvolgyi Ut 72C Budapest (XII), Hungary
Fairfield University, Fairfield Connecticut 06430
University of Connecticut, Waterbury, Connecticut 06710