ON A COMBINATORIAL PROBLEM

P. ERDŐS

Let $\mathcal{F}$ be a family of sets. $\mathcal{F}$ is said by E. W. Miller [3] to possess property B if there exists a set $B$ such that

- $F \cap B \neq \emptyset$ for every $F \in \mathcal{F}$,
- $F \cap B = \emptyset$ for every $F \in \mathcal{F}$.

Miller used the letter B in honour of Felix Bernstein, who in the early years of this century proved that the perfect sets have property B and using this "constructed" a totally imperfect set of power continuum (that is, a set of power continuum which does not contain a perfect set). I put constructed in quotation mark, since he used the axiom of choice (in fact, without the axiom of choice the existence of a totally imperfect set has never been proved).

Several other well known theorems can be formulated in terms of property B. For example, a well known theorem of van der Waerden states that if we split the integers into two classes, then at least one class contains for every $k$ an arithmetic progression of $k$ terms. This theorem can be formulated as follows: The family of all arithmetic progressions of $k$ terms does not have property B.

Hajnal and I [2] recently published a paper on the property B and its generalizations. One of the unsolved problems we state asks: What is the smallest integer $m(p)$ for which there exists a family $\mathcal{F}$ of finite sets $A_1, \ldots, A_{m(p)}$, each having $p$ elements, which does not possess property B?

For $p = 1$ there is no problem since $m(1) = 1$. Trivially $m(2) = 3$ and by trial and error we showed $m(3) = 7$. $m(3) \leq 7$ is shown by the set of Steiner triplets $(1,2,3), (1,4,5), (1,6,7), (2,4,7), (2,5,6), (3,4,6), (3,5,7)$. It is easy to see that every set which has a non-empty intersection with each of these sets must contain at least one of them. By a somewhat longer trial and error method we showed $m(3) > 6$. Thus $m(3) = 7$. The value of $m(p)$ is not known for $p > 3$ and it does not seem easy to determine $m(p)$ even for $p = 4$. We further observed that $m(p) \leq \binom{2p-1}{p}$ by
defining the family $\mathfrak{F}$ as the set of all subsets taken $p$ at a time of a set of $2p - 1$ elements.

We shall now show that for all $p \geq 2$:

1. \[ m(p) > 2^{p-1}, \]

and for every $\varepsilon > 0$ if $p > p_0(\varepsilon)$:

2. \[ m(p) > (1 - \varepsilon)2^p \log 2. \]

$\overline{A}_k$ will denote the number of elements of $A_k$, and $A_i \setminus A_j$ will denote the set of those elements of $A_i$ which are not contained in $A_j$. Instead of (1) and (2) we shall prove the following

**Theorem 1.** Let $\{A_i\}$, $1 \leq i \leq k$ be a family $\mathfrak{F}$ of finite sets, $\overline{A}_i = \alpha_i \geq 2$. If

3. \[ \sum_{i=1}^{k} \frac{1}{2^{\alpha_i}} \leq \frac{1}{2}, \]

or

4. \[ \prod_{i=1}^{k} \left(1 - \frac{1}{2^{\alpha_i}}\right) \geq \frac{1}{2}, \]

holds, then $\mathfrak{F}$ has property $B$.

(1) clearly follows from (3) and (2) from (4). In fact (4) clearly implies (3), and we include (3) only because its proof is very simple.

I do not know the order of magnitude of $m(p)$ and cannot even prove that

5. \[ \lim_{p \to \infty} m(p)^{1/p} \]

exists. Quite possibly the limit in (5) is 2.

Put $\bigcup_{i=1}^{k} A_i = T$, $T = n$. If $\mathfrak{F}$ is a family of sets, $\overline{\mathfrak{F}}$ will denote the number of sets in the family. Denote by $\overline{\mathfrak{F}}_T$ the family of sets $S$ for which

6. \[ S \subseteq T, A_i \cap S \neq \emptyset, A_i \in S, 1 \leq i \leq k. \]

We have to show that if (3) holds then $\overline{\mathfrak{F}}_T > 0$ (since this implies that the family of sets $A_i$, $1 \leq i \leq k$ satisfying (3) has property $B$). Denote by $\overline{\mathfrak{F}}_i$ the family of sets $S$ satisfying

7. \[ S \subseteq T, A_i \subseteq S \quad \text{or} \quad A_i \cap S = \emptyset. \]

Clearly an $S \subseteq T$ is in the family $\overline{\mathfrak{F}}_T$ if it is in none of the families $\overline{\mathfrak{F}}_i$, $1 \leq i \leq k$ (that is, it satisfies (6) if it does not satisfy (7) for any $i$, $1 \leq i \leq k$). By a simple sieve process we thus have

8. \[ \overline{\mathfrak{F}}_T \geq 2^n - \sum_{i=1}^{k} \overline{\mathfrak{F}}_i + 1. \]
The proof of (8) is indeed easy. \(2^n\) is the number of all subsets of \(T\), and to obtain \(\overline{\Omega}_T\) we have to subtract away all the sets of \(\overline{\Omega}_i\), \(1 \leq i \leq k\). But the sets which contain \(A_1 \cup A_2\) have been subtracted away twice and there is at least one such set (namely \(T\)), which explains the summand \(+1\) on the right hand side of (8). We evidently have

\[
\overline{\Omega}_i = 2^{n-n_i+1},
\]

since clearly there are \(2^{n-n_i}\) sets \(S \subseteq T\) satisfying \(A_i \subseteq S\) and \(2^{n-n_i}\) sets satisfying \(A_i \cap S = \emptyset\). From (8) and (9) we have \(\overline{\Omega}_T \geq 1\) if (3) is satisfied. This proves the first statement of Theorem 1.

To prove the second statement we need the following

**Lemma.** Let \(T \subseteq T_1\), \(T_1 = m \geq n\). The number of subsets \(S \subseteq T_1\) which do not contain any of the sets \(A_i\), \(1 \leq i \leq k\) is greater than or equal to

\[
\frac{2m}{k} \left(1 - \frac{1}{2^{n_i}}\right),
\]

with equality if and only if the sets \(A_i\) are pairwise disjoint.

We use the set \(T_1 \supseteq T\) only to make our induction proof easier. Denote by \(f(A_1, \ldots, A_j; T_1)\) the number of subsets \(S\) of \(T_1\) not containing any of the sets \(A_i\), \(1 \leq i \leq j\), and by \(f(A_1, \ldots, A_j; A_{j+1}, T_1)\) the number of sets \(S \subseteq T_1\) which contain \(A_{j+1}\), but do not contain any of the sets \(A_i\), \(1 \leq i \leq j\).

If the sets \(A_i\) are pairwise disjoint, we evidently have

\[
f(A_1, \ldots, A_k; T_1) = 2^{m-n} \prod_{i=1}^k \left(2^{n_i} - 1\right) = 2^m \prod_{i=1}^k \left(1 - \frac{1}{2^{n_i}}\right),
\]

since we obtain the sets \(S \subseteq T_1\), \(A_i \not\subseteq S\), \(1 \leq i \leq k\) by taking the unions of all the proper subsets of the sets \(A_i\) with any subset of \(T_1 \setminus T\). Thus there is equality in (10).

Assume next that the sets \(A_i\) are not pairwise disjoint, say \(A_i \cap A_j \neq \emptyset\). If \(k = 2\), a simple argument shows that

\[
f(A_1, A_2; T_1) = 2^m - 2^{m-n} - 2^m - 2^{m-n} > 2^m \left(1 - \frac{1}{2^{n_1}}\right) \left(1 - \frac{1}{2^{n_2}}\right),
\]

where \(n = A_1 \cup A_2 \leq n_1 + n_2\). Thus for \(k = 2\) (10) holds with the sign of inequality. Assume next that if we have any \(k-1\) \((k \geq 3)\) sets which are not pairwise disjoint, then (10) holds with the sign of inequality. We shall show that the same is true for \(k\) sets \(A_1, \ldots, A_k\), \(A_1 \cap A_2 \neq \emptyset\).

By a simple argument we have

\[
f(A_1, \ldots, A_k; T_1) = f(A_1, \ldots, A_{k-1}; T_1) - f(A_1, \ldots, A_{k-1}; A_k, T_1).
\]
By our induction hypothesis we have
\[(13) \quad f(A_1, \ldots, A_{k-1}; T_1) > 2^m \prod_{i=1}^{k-1} \left(1 - \frac{1}{2^{a_i}}\right).\]

Further clearly
\[(14) \quad f(A_1, \ldots, A_{k-1}; A_k, T_1) = f(A_1 \setminus A_k, \ldots, A_{k-1} \setminus A_k; T_1 \setminus A_k).\]

To every subset \(S'\) of \(T_1 \setminus A_k\) which does not contain any of the sets \(A_i \setminus A_k, 1 \leq i \leq k-1\), we make correspond \(2^{2a_i}\) subsets \(S\) of \(T_1\) which do not contain any of the sets \(A_i, 1 \leq i \leq k-1\). It suffices to consider the sets
\[(15) \quad S' \cup S'', \quad S'' \subseteq A_k.\]

Clearly if two subsets \(S'_1\) and \(S'_2\) of \(T_1 \setminus A_k\) are distinct, all the sets \(15\) are distinct. Thus we have
\[(16) \quad f(A_1 \setminus A_k, \ldots, A_{k-1} \setminus A_k; T_1 \setminus A_k) \leq \frac{f(A_1, \ldots, A_{k-1}; T_1)}{2^{2a_i}}.\]

From (12), (13), (14), and (16) we obtain
\[f(A_1, \ldots, A_k; T_1) \geq f(A_1, \ldots, A_{k-1}; T_1) \left(1 - \frac{1}{2^{a_i}}\right) > 2^m \prod_{i=1}^{k} \left(1 - \frac{1}{2^{a_i}}\right),\]
which proves the Lemma.

The Lemma in fact follows immediately from the following special case of a theorem of Chung [1]: Let \(E_i, 1 \leq i \leq k\) be \(k\) events of probability \(\beta_i\), \(E_i^c\) denoting the event (of probability \(1 - \beta_i\)) that \(E_i\) does not happen. Assume that for every \(i, 2 \leq i \leq k:\)
\[(17) \quad P(E_1 \cup \ldots \cup E_{i-1} | E_i^c) \geq P(E_1 \cup \ldots \cup E_{i-1}),\]
where \(P(E | F)\) denotes the conditional probability of \(E\) happening if we know that \(F\) has happened. (17) implies
\[(18) \quad P(E_1 \cap \ldots \cap E_k) \geq \prod_{i=1}^{k} (1 - \beta_i),\]
with equality only if there is equality in (17) for every \(i, 2 \leq i \leq k\). We obtain our Lemma by defining the event \(E_i^c\) as the event that \(S \subseteq T_1\) contains \(A_i\).

To complete the proof of Theorem 1 we have to show that if \(A_i, 1 \leq i \leq k\) satisfies (4), then \(\overrightarrow{\gamma}_T > 0\) (see the proof of (3)). Clearly
\[2^a - f(A_1, \ldots, A_k; T)\]
equalsthe number of subsets \(S\) of \(T\) for which \(A_i \subseteq S\) holds for some \(i, 1 \leq i \leq k\), and it also equals the number of subsets
ON A COMBINATORIAL PROBLEM

S ⊂ T for which \( A_i \subset T \setminus S \) for some \( i, 1 \leq i \leq k \). Denote by \( L \) the number of subsets \( S \subset T \) for which \( A_{i_1} \subset S \) and \( A_{i_2} \subset T \setminus S \) for some \( i_1 \) and \( i_2 \).

A simple argument shows that

\[
\mathcal{F}_T = 2^n - 2^{2n - f(A_1, \ldots, A_k; T)} + L.
\]

If the sets \( A_i \) are not pairwise disjoint, then (4), (19) and our Lemma implies \( \mathcal{F}_T > 0 \). If the sets \( A_i \) are pairwise disjoint (in fact if \( A_1 \cap A_2 = \emptyset \)), then \( L > 0 \) since \( A_1 \subset A_1, A_2 \subset T - A_1 \). Thus in any case (4) implies \( \mathcal{F}_T > 0 \) and hence Theorem 1 is proved.

By slightly more complicated arguments we could prove the following

**Theorem 2.** Let \( A_1, A_2, \ldots \) be a finite or infinite sequence of finite sets satisfying

\[
\mathcal{A}_i \geq 2 \quad \text{and} \quad \prod_i \left( 1 - \frac{1}{2^{a_i}} \right) \geq \frac{1}{2},
\]

and \( A'_1, A'_2, \ldots \) a finite or infinite sequence of infinite sets. Then the family \( \{A_i\} \cup \{A'_i\} \) has property B.

Now one can ask the following problem which I cannot answer: Let \( \{A_i\} \) be a finite or infinite family of finite sets which does not have the property B and for which \( \overline{A}_i \geq p \geq 2 \) for all \( i \). What is the upper bound \( C^{(p)} \) of \( \prod_i (1 - 2^{-a_i}) \) and the lower bound \( C_p \) of \( \sum_i 2^{-a_i} \)? Very likely \( C^{(p)} = \frac{2^p}{4} \) and \( C_2 = \frac{3}{4} \). Probably

\[
\lim_{p \to \infty} C^{(p)} = 0, \quad \lim_{p \to \infty} C_p = \infty.
\]

If \( f(A_1, \ldots, A_k; T) > 2^{n-1} \), our proof immediately shows that the family \( \{A_i\}, 1 \leq i \leq k \) has property B, but if \( f(A_1, \ldots, A_k; T) = 2^{n-1} \), the family \( \{A_i\}, 1 \leq i \leq k \) does not have to have property B, for instance if it consists of the subsets taken \( p \) at a time of a set of \( 2p - 1 \) elements.

A family of sets \( \mathcal{F} \) is said to have property B(s) if there exists a set \( B \) such that \( F \cap B \neq \emptyset \) and \( F \cap \overline{B} < s \) for every \( F \) of \( \mathcal{F} \).

Hajnal and I asked [2] what is the smallest integer \( m(p, s) \) for which there exists a family \( \mathcal{F} \) of sets \( A_i, 1 \leq i \leq m(p, s) \) not having property B(s) and satisfying \( \overline{A}_i = p, 1 \leq i \leq m(p, s) \). Clearly \( m(p, p) = m(p) \), and we remarked that \( m(p, s) \leq \left( \frac{p + s + 1}{s} \right) \).

Using the methods of this note we can show that positive absolute constants \( c_1 \) and \( c_2 \) exist so that

\[
(1 + c_1)^s < m(p, s) < (1 + c_2)^s.
\]
REFERENCES

