SIAM J. ALG. DISC. METH.
Vol. 1, No. 4, December 1980

0196-5212/80/0104-0003 $01.00/0

THE ERDÖS-KO-RADO THEOREM FOR INTEGER SEQUENCES*

PETER FRANKL† AND ZOLTÁN FÜREDI‡

Abstract. For positive integers n, k, t we investigate the problem how many integer sequences $(a_1, a_2, \cdots, a_n)$ we can take, such that $1 \leqq a_i \leqq k$ for $1 \leqq i \leqq n$, and any two sequences agree in at least t positions. This problem was solved by Kleitman (J. Combin. Theory, 1 (1966), pp. 209–214) for $k = 2$, and by Berge (in *Hypergraph Seminar, Columbus, Ohio* (1972), Springer-Verlag, New York, 1974) for $t = 1$. We prove that for $t \geqq 15$ the maximum number of such sequences is k^{n-t} if and only if $k \geqq t+1$.

1. Introduction. Let t, k, n be positive integers with $k \geqq 2$, $n \geqq t$, and let $\mathscr{A}$ be a set of integer sequences $(a_1, a_2, \cdots, a_n)$, $1 \leqq a_i \leqq k$. We say that $(a_1, a_2, \cdots, a_n)$ and $(a'_1, a'_2, \cdots, a'_n)$ intersect in at least t positions if we can find $1 \leqq i_1 < i_2 < \cdots < i_t \leqq n$ such that $a_{i_j} = a'_{i_j}$ for $j = 1, \cdots, t$.

Let $f(n, k, t)$ denote the maximum cardinality $\mathscr{A}$ can have supposing that any two elements of $\mathscr{A}$ intersect in at least t positions. Setting $\mathscr{A}_0 = \{(a_1, \cdots, a_n) | 1 \leqq a_i \leqq k, a_j = 1 \text{ for } j = 1, \cdots, t\}$, we obtain

$$f(n, k, t) \geqq k^{n-t}. \tag{1}$$

In the case $k = 2$ the problem reduces to the following: What is the maximum number of subsets of an n-set such that the symmetric difference of any two has cardinality at most $n - t$? This problem was posed by Erdös and solved by Kleitman [5], who proved that

$$f(n, 2, t) = \begin{cases} \sum_{i=0}^{[(n-t)/2]} \binom{n}{i} & \text{if } n-t \text{ is even,} \\ 2\sum_{i=0}^{[(n-t)/2]} \binom{n-1}{i} & \text{if } n-t \text{ is odd.} \end{cases} \tag{2}$$

The expression (2) is much greater than (1) except for $t = 1$, when we have equality. Berge [1] proved that

$$f(n, k, 1) = k^{n-1} \tag{3}$$

holds for $k \geqq 3$ as well. In fact he proved that if instead of $a_i \leqq k$ we suppose $a_i \leqq k_i$, $k_1 \leqq \cdots \leqq k_n$, then the corresponding bound is $k_2 k_3 \cdots k_n$. Livingston [7] proved that if equality holds in (3) then necessarily $\mathscr{A}$ is of the form $\mathscr{A}_0$ (up to isomorphism). In the present paper we are mainly concerned with the problem, for which triples n, k, t is the bound (1) optimal. We have the following

CONJECTURE. *The bound* (1) *is optimal if and only if* $n \leqq t+1$ *or* $k \geqq t+1$.

Remark. It is easy to check that the conjecture holds for $n \leqq t+1$, i.e., $n = t$ and $n = t+1$. On the other hand, (2) and (3) settle it for $t = 1$.

THEOREM 1. *The conjecture holds for* $t \geqq 15$.

We give some results for the range $2 \leqq t \leqq 14$ as well.

2. Preliminaries. Our main tool in proving Theorem 1 will be the strongest form of the Erdös-Ko-Rado theorem (see [2]), proved in Frankl [3]. To state it we need some

* Received by the editors December 3, 1979, and in final form March 24, 1980.

† Centre National de la Recherche Scientifique, Paris, France.
‡ Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary.

definitions. Let s be an integer $t \leqq s \leqq n$. Let $\mathscr{B}$ be a family of s-element subsets of $\{1, 2, \cdots, n\}$ satisfying $|B \cap B'| \geqq t$ for $B, B' \in \mathscr{B}$. Such a family is called t-intersecting. Let us define

$$\mathscr{B}_0 = \{B \subseteq \{1, 2, \cdots, n\} \mid \{1, 2, \cdots, t\} \subseteq B, |B| = s\}$$

$$\mathscr{B}_1 = \{B \subseteq \{1, 2, \cdots, n\} \mid |\{1, 2, \cdots, t+2\} \cap B| \geqq t+1, |B| = s\}.$$

Clearly both $\mathscr{B}_0$ and $\mathscr{B}_1$ are t-intersecting. Then we have

THEOREM 2[3]. *There exist positive constants c_t, depending on t only and satisfying $c_t < 2$ for $t \geqq 2$, and $c_t < 1$ for $t \geqq 15$ such that for*

$$\frac{n-t}{s-t} > c_t(t+1), \tag{4}$$

a t-intersecting family of maximum size is of the form $\mathscr{B}_0$ or $\mathscr{B}_1$, (up to isomorphism).

As remarked in [3],

$$|\mathscr{B}_0| \lesseqqgtr |\mathscr{B}_1| \quad \text{iff} \quad n \gtreqqless (s-t+1)(t+1). \tag{5}$$

Let us now return to our set of sequences $\mathscr{A}$, which is t-intersecting; i.e., any two sequences in $\mathscr{A}$ intersect in at least t positions. Let us define:

$$\mathscr{A}^+ = \{(a_1, a_2, \cdots, a_{n+1}) \mid (a_1, \cdots, a_n) \in \mathscr{A}, 1 \leqq a_{n+1} \leqq k\}.$$

It is evident that $\mathscr{A}^+$ is t-intersecting, yielding

$$f(n+1, k, t) \geqq kf(n, k, t). \tag{6}$$

Consequently the function $f(n, k, t)k^{-n}$ is nondecreasing in n (and bounded by 1). Hence the following limit exists (and is at most 1):

$$g(k, t) = \lim_{n \to \infty} f(n, k)k^{-n}.$$

We will now bring $\mathscr{A}$ to a canonical form. Such a transformation was first used by Kleitman [6]. Let $1 \leqq j \leqq n$. Define the transformation,

$$T_{j,i}(a_1, \cdots, a_j \cdots, a_n) = \begin{cases} (a_1, a_2, \cdots, a_j', \cdots, a_n) & \text{if this sequence is not in } \mathscr{A}, \text{ and} \\ & a_j = i,\ a_j' = 1; \\ (a_1, a_2, \cdots, a_j, \cdots, a_n) & \text{otherwise.} \end{cases}$$

It is easily seen that $T_{j,i}(\mathscr{A}) = \{T_{j,i}(A) \mid A \in \mathscr{A}\}$ is t-intersecting and has the same cardinality as $\mathscr{A}$. Repeated application of the transformation yields a system $\mathscr{A}'$ which is t-intersecting, $|\mathscr{A}'| = |\mathscr{A}|$, and for $1 \leqq j \leqq n$, $1 \leqq i \leqq k$,

$$T_{j,i}(\mathscr{A}') = \mathscr{A}'. \tag{7}$$

Without loss of generality we may assume $\mathscr{A} = \mathscr{A}'$. Let us associate with every $(a_1, \cdots, a_n) = A$, the set $B(A) = \{i \mid a_i = 1\}$.

PROPOSITION 1. *The family $\mathscr{B} = \{B(A) \mid A \in \mathscr{A}\}$ is t-intersecting.*

Proof. Let $A = (a_1, \cdots, a_n)$, $A' = (a_1', \cdots, a_n') \in \mathscr{A}$. Let $\{i_1, i_2, \cdots, i_r\}$ be the set of i's such that $a_i = a_i' \neq 1$. In view of (7), $A'' = (a_{1'}'', \cdots, a_n'') \in \mathscr{A}$, where $a_i'' = a_i'$ for $i \notin \{i_1, \cdots, i_r\}$, $a_i'' = 1$ for $i \in \{i_1, \cdots, i_r\}$. As $(a_1, \cdots, a_n)$ and $(a_{1'}'', \cdots, a_n'')$ agree in the ith position only for $i \in B(A) \cap B(A')$, the statement of the proposition follows. Now by the maximality of $\mathscr{A}$ we have

PROPOSITION 2. $\mathscr{A} = \{A = (a_1, \cdots, a_n) \mid 1 \leqq a_i \leqq k, B(A) \in \mathscr{B}\}$, *and consequently*

$$|\mathscr{A}| = \sum_{B \in \mathscr{B}} (k-1)^{n-|B|}. \tag{8}$$

Hence the problem of determining $f(n, k, t)$ reduces to finding the maximum of (8) over all t-intersecting families $\mathcal{B}$. We need an easy probabilistic result.

PROPOSITION 3. *For every positive ε and δ the number of sequences $(a_1, \cdots, a_n)$ with $1 \leqq a_i \leqq k$ which contain more than $(1+\varepsilon)(n/k)$ 1's or less than $(1-\varepsilon)(n/k)$ 1's is less than δk^n for $n > n_0(\delta, \varepsilon)$.*

Instead of a proof, just observe that $p(a_i = 1) = 1/k$; hence the mean value of 1's is n/k, and the events $a_i = 1$ are independent for $i = 1, \cdots, n$.

3. The main results. We prove Theorem 1 in a somewhat surprising way; namely we prove first that it holds asymptotically, i.e., $f(n, k, t) \leqq (1+o(1))k^{n-t}$ for k, t fixed, $k > t \geqq 15$. Then we deduce $f(n, k, t) = k^{n-t}$ from it for every $n \geqq t$.

THEOREM 3. *For $k > t \geqq 15$ we have*

$$g(k, t) = \lim_{n\to\infty} f(n, k, t)k^{-n} = k^{-t}.$$

In view of Proposition 2,

$$f(n, k, t)k^{-n} = \left(\sum_{B\in\mathcal{B}} (k-1)^{n-|B|}\right)k^{-n},$$

for some t-intersecting family $\mathcal{B}$. Moreover, Proposition 3 gives that for any $\delta, \varepsilon > 0$, $n > n_0(\delta, \varepsilon)$, we have

$$f(n, k, t)k^{-n} < \left(\sum_{B} (k-1)^{n-|B|}\right)k^{-n} + \delta, \tag{9}$$

where B runs over those elements of $\mathcal{B}$ which satisfy

$$(1-\varepsilon)(n/k) \leqq |B| \leqq (1+\varepsilon)(n/k).$$

Now for $(1-\varepsilon)(n/k) \leqq s \leqq (1+\varepsilon)(n/k)$, set

$$\mathcal{B}(s) = \{B \in \mathcal{B} \,|\, |B| = s\}.$$

As $k \geqq t+1$, for $n > n_0(\varepsilon)$ we have $(n-t)/(s-t) > c_t(t+1)$; i.e., (4) is satisfied and we may apply Theorem 2 to the t-intersecting family $\mathcal{B}(s)$. We deduce

$$|\mathcal{B}(s)| \leqq \max(|\mathcal{B}_0|, |\mathcal{B}_1|) = \max\left(\binom{n-t}{s-t}, (t+2)\binom{n-t-2}{s-t-1} + \binom{n-t-2}{s-t-1}\right). \tag{10}$$

By (5) for $k > t+1$ the value of (10) is $\binom{n-t}{s-t}$.

If $k = t+1$, then

$$\frac{1-\varepsilon}{t+1} < \frac{s}{n} < \frac{1+\varepsilon}{t+1}.$$

We can still obtain for $n > n_0(t, \varepsilon)$,

$$\binom{n-t-2}{s-t-1}\Big/\binom{n-t}{s-t} = \frac{(s-t)\cdot(n-s)}{(n-t)\cdot(n-t+1)} < \frac{(t+\varepsilon)(1+\varepsilon)}{t^2+2t+1-\varepsilon},$$

and

$$\binom{n-t-s}{s-t-2}\Big/\binom{n-t}{s-t} = \frac{s-t}{n-t}\cdot\frac{s-t-1}{n-t-1} < \frac{(1+\varepsilon)^2}{t^2+2t+1},$$

yielding

$$|\mathscr{B}(s)| \leqq |\mathscr{B}_0| \frac{(t+2)(t+\varepsilon)(1+\varepsilon)+(1+\varepsilon)^2}{t^2+2t+1-\varepsilon} < \binom{n-t}{s-t}(1+2\varepsilon), \tag{11}$$

whenever ε is sufficiently small. Now from (9) and (11) we obtain

$$\begin{aligned} f(n, k, t)k^{-n} &< (1+2\varepsilon) \sum_{s=t}^{n} \binom{n-t}{s-t}(k-1)^{n-s}k^{-n} + \delta \\ &= (1+2\varepsilon)k^{-n} \sum_{j=0}^{n-t} \binom{n-t}{j}(k-1)^j + \delta \\ &= (1+2\varepsilon)k^{-n}k^{n-t} + \delta \\ &= (1+2\varepsilon)k^{-t} + \delta, \end{aligned}$$

which implies, since $\delta, \varepsilon > 0$ were arbitrary,

$$g(k, t) \leqq k^{-t}.$$

As $|\mathscr{A}_0| = k^{n-t}$ we have $g(k, t) \geqq k^{-t}$ as well, which concludes the proof of Theorem 3.

Proof of Theorem 1. Suppose that for some t-intersecting family $\mathscr{A}$ we have $|\mathscr{A}| \geqq k^{n-t}+1$. Then using (6) we deduce

$$f(n', k, t) \geqq k^{n'-n} f(n, k, t) \geqq k^{n'-n}|\mathscr{A}| \geqq k^{n'}(k^{-t}+k^{-n}),$$

whence $g(k, t) \geqq k^{-t} + k^{-n} > k^{-t}$, a contradiction (observe that now n is fixed and we have $n' \to \infty$), which proves the *if* part of Theorem 1.

For the *only if* part, let us define

$$\mathscr{A}_1 = \{A = (a_1, \cdots, a_n) \,|\, 1 \leqq a_i \leqq k, |B(\mathscr{A}) \cap \{1, \cdots, t+2\}| \geqq t+1\}.$$

Obviously $\mathscr{A}_1$ is t-intersecting, and we have for $n \geqq t+2$, $k \leqq t$,

$$|\mathscr{A}_1| = k^{n-t-2}((t+2)(k-1)+1) = k^{n-t}(1+(t+1-k)k^{-2}) > k^{n-t}.$$

4. Some remarks. Using the same argument we could deduce

THEOREM 4. *If* $t \geqq k > c_t(t+1)$, *then*

$$g(k, t) = k^{-t-2}((t+1)(k-1)+1).$$

(By [3] we know that $c_t < 0.8$ for $t \geqq 15$.) Now Theorem 4 yields

THEOREM 5. *If* $t \geqq k > c_t(t+1)$, *then*

$$f(n, k, t) = k^{n-t-2}((t+1)(k-1)+1).$$

5. Probabilistic arguments. Now we want to apply the random walk method developed in [4] to obtain a general bound on $g(k, t)$, $k > 2$.

Let $\mathscr{B}$ be the t-intersecting family associated with the maximal set of t-intersecting sequences $\mathscr{A}$. With $\mathscr{B}$ we proceed as in [3]. For $1 \leqq i < j \leqq n$, the canonical transformation is the following.

$$K_{i,j}(B) = \begin{cases} B' = B - \{j\} \cup \{i\} & \text{if } i \notin B, j \in B, B' \notin \mathscr{B}, \\ B & \text{otherwise}; \end{cases}$$

$$K_{i,j}(\mathscr{B}) = \{K_{i,j}(B) \,|\, B \in \mathscr{B}\}.$$

Applying $K_{i,j}$ repeatedly we obtain a t-intersecting family $\mathscr{B}'$ which satisfies $K_{i,j}(\mathscr{B}')=\mathscr{B}'$ for all $1 \leqq i<j \leqq n$. We may suppose $\mathscr{B}=\mathscr{B}'$. The following propositions are taken essentially from [3].

PROPOSITION 4. *No subset S of $T=\{1, 2, \cdots, t-1, t+1, \cdots, t+2l+1, \cdots\}$ belongs to $\mathscr{B}$.*

Proof. Otherwise an application of $K_{t+2l,t+2l+1}$ for all $l \geqq 0$ would yield $S' \in \mathscr{B}$ for $S' \subseteq T'=\{1, 2, \cdots, t, t+2, \cdots, t+2l, \cdots\}$. But $|S \cap S'| \leqq |T \cap T'|=t-1$, a contradiction.

Let us associate with a sequence $A=(a_1, \cdots, a_n)$ a random walk in the plane in the following way. We start from $(0, 0)$. Suppose that after $(i-1)$ moves we are in (x, y). Then we move to $(x, y+1)$ or $(x+1, y)$ according to whether $a_i=1$ or not. The random walks associated with different sets are different. Proposition 4 yields (see [3])

PROPOSITION 5. *The random walk associated with $A \in \mathscr{A}$ hits the line $y=x+t$.*

In probability language, considering the space of all possible sequences $(a_1, \cdots, a_n)$, $1 \leqq a_i \leqq k$, we move upward with probability $1/k$ and to the right with probability $(k-1)/k$. Now let us continue to walk indefinitely. Then for the probability of hitting the line $y=x+t$, $p(t)$ we obtain

$$p(0)=1,$$

$$p(t)=(1/k)p(t-1)+((k-1)/k)p(t+1) \quad \text{for } t \geqq 1,$$

and

$$p(t) \to 0 \quad \text{as } t \to \infty, \text{ because } k>2.$$

Hence, we deduce

$$p(t)=(k-1)^{-t}.$$

Consequently we have

THEOREM 6. *For $k \geqq 3$ we have $k^{-n}f(n, k, t) \leqq (k-1)^{-t}$, and consequently*

$$g(k, t) \leqq (k-1)^{-t}. \tag{12}$$

From (2) it follows that $g(2, t)=2^{-1}$ for every $t \geqq 1$, which is a great contrast to (12).

On the other hand, for k, t fixed, let $(s, s+t)$ be the point of the line $y=x+t$ for which the probability that a random walk goes through it is the largest. Let $\mathscr{A}_s$ be the set of the corresponding sequences. Then obviously

$$\mathscr{A}_s=\{A=(a_1, \cdots, a_n) \,|\, |B(A) \cap \{1, 2, \cdots, t+2s\}| \geqq t+s\}. \tag{13}$$

Thus

$$g(k, t) \geqq \frac{|A_s|}{k^n}.$$

Then elementary computation shows that for some constant d_k depending on k only, we have

$$g(k, t) > \frac{(k-1)^{-t}}{d_k \sqrt{t}}.$$

Let us finish with a conjecture, setting the general case.

CONJECTURE. *Let $\mathscr{A}_s$ be defined by* (13). *Then*

$$f(n, k, t)=\max_{s \geqq 0} |\mathscr{A}_s|.$$

REFERENCES

[1] C. Berge, *Nombres de coloration de l'hypergraphe h-parti complet*, Hypergraph seminar, Columbus, Ohio, 1972, Springer-Verlag, New York, 1974, pp. 13–20.

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

[3] P. Frankl, *The Erdös-Ko-Rado theorem is true for* $n = ckt$, Proc. Fifth Hung. Comb. Coll. Keszthely, 1976, North-Holland, Amsterdam, 1978, pp. 365–375.

[4] ———, *Families of finite sets satisfying an intersection condition*, Bull. Austral. Math. Soc., 15 (1976), pp. 73–79.

[5] D. J. Kleitman, *On a combinatorial conjecture of Erdös*, J. Combin. Theory, 1 (1966), pp. 209–214.

[6] ———, *Families of non-disjoint subsets*, J. Combin. Theory, 1 (1966), pp. 153–155.

[7] M. L. Livingston, *An ordered version of the Erdös-Ko-Rado theorem*, J. Combin. Theory, Ser. A, 26 (1979), pp. 162–165.