Journal of Statistical Planning and Inference 7 (1982) 139-150
North-Holland Publishing Company

SEQUENTIAL SEARCH USING QUESTION-SETS WITH BOUNDED INTERSECTIONS

András SEBŐ

Computer and Automation Institute, Hungarian Academy of Sciences, Budapest, Hungary

Received 15 June 1979; revised manuscript received 20 September 1980
Recommended by G.O.H. Katona

Abstract: The paper is concerned with sequential search on a finite set: an unknown element of the finite set is to be found testing its subsets, and getting the information that the unknown element is or is not an element of the tested subset. The optimum of the strategy lengths is found under the condition that the intersection of any two different test-sets is bounded. This condition is generalized taking the intersection of any m different test-sets instead of two, and the general problem is also solved.

AMS Subject Classification: 94A50, 05B99.

Key words: Search problem; Sequential and static strategies; Separating systems.

1. Introduction

We first introduce some notations: $|A|$ is the cardinality of the set A; H is a given set with $n=|H|$, $\mathbb{Z}$ denotes the set of all natural numbers; for any real number $\lfloor u \rfloor$ denotes the unique integer satisfying $u-1<\lfloor u \rfloor \leq u$, and $\lceil u \rceil$ the unique integer satisfying $u \leq \lceil u \rceil < u+1$; $k \mid n$ means that n is a multiple of k and $k \nmid n$ means that n is not a multiple of k; $(a, b]$ denotes the set of all real numbers $a<z\leq b$; log is always understood as base 2 logarithm; $\emptyset$ is the empty set.

The paper investigates the search problem with one unknown $x \in H$ to be found using subsets $A_1, A_2(x), \ldots$ of H as questions $(A_i(x)=A_i(t_1(x), \ldots, t_{i-1}(x)))$. $A_i(x)$ may depend on the answers $t_j(x)$ to the previous questions $A_j(x)$ $(j=1,2,\ldots,i-1)$, $t_j(x)=1$ if $x \in A_j(x)$ and $t_j(x)=0$ if $x \notin A_j(x)$. $A_i(x)$ will be called *question-set*. A rule which defines $A_i(x)$ in knowledge of $t_1(x), \ldots, t_{i-1}(x)$ is called *strategy*. Denote by Σ the set of all strategies. Let the search *length* $N(x)$ be defined as the smallest integer r such that

$$t_1(x), t_2(x), \ldots, t_r(x)$$

uniquely specifies x. We write sometimes $N_S(x)$ in order to emphasize that it belongs to the strategy S. If for all i, $A_i(x)$ is independent of the value of $t_0(x), t_1(x), \ldots, t_{i-1}(x)$ then the strategy is called *static*, otherwise *sequential*. A strategy S_0 is called *optimal* (in min–max sense) if

$$\max_{x\in H} N_{S_0}(x) = \min_{S\in\Sigma} \max_{x\in H} N_S(x).$$

$N_S = \max_{x\in H} N_S(x)$ is called the length of the strategy S.

Remark. Our search model is the model of the braingame 20 questions (also called Bar-Kochba game). It is also to be noted that static strategies correspond to 'systems of separating sets' (see Katona (1973)).

It is evident that without more restrictions the optimal strategy length is

$$\lceil \log n \rceil$$

independently of whether the search is sequential or static.

A. Rényi has posed the problem of finding the length of the optimal strategy if the question-sets are required to be less than or equal to some natural number k fixed in advance. This problem was solved by G.O.H. Katona, for both sequential and static searches (Katona (1966) and (1973)).

Katona has proposed the condition that the intersection of any two question-sets is less than or equal to some pre-fixed number k. His investigations concern the static search (Katona (1976)).

In this paper the corresponding sequential search problem is solved: two sequential search problems seem to arise, the first with the condition $|A_i(x)\cap A_j(y)|\le k$ for all $x, y\in H$, $1\le i$, $j\le N$, $A_i(x)\ne A_j(y)$, and another if we require $|A_i(x)\cap A_j(x)|\le k$ for all $x\in H$, $1\le i$, $j\le N$, $i\ne j$. It is easy to see that the two conditions are equivalent. (If the answer for a question A is "yes" then the following questions can be supposed to be subsets of A (if they are not, their intersection with A can be taken), and if it is "no", then they can be supposed to be subsets of $H\setminus A$. Such a strategy has the following property: for all pairs $A_i(x), A_j(y)$ either $A_i(x)=A_j(y)$ or $A_i(x)\cap A_j(y)=\emptyset$.)

Let $f_1(n,k)$ denote the minimal length of the strategies satisfying

$$|A_i(x)|\le k \quad \text{for all } x\in H,\ 1\le i\le N, \tag{1}$$

and $f_2(n,k)$ the minimal length of the strategies satisfying

$$|A_i(x)\cap A_j(x)|\le k \quad \text{for all } x\in H,\ 1\le i\le j\le N. \tag{2}$$

These conditions are generalized to

$$|A_{i_1}(x)\cap A_{i_2}(x)\cap\cdots\cap A_{i_m}(x)|\le k \quad \text{for all } x\in H,\ 1\le i_1<\cdots<i_m\le N. \tag{m}$$

where $m\in\mathbb{Z}$ is fixed arbitrarily.

Let $f_m(n,k)$ denote the minimal length of the strategies satisfying the condition (m).

We will find the exact value of $f_2(n,k)$ and recursive formulas will be proved for the generalization allowing the calculation of $f_m(n,k)$ and also the construction of the optimal strategies for all m, n, k.

2. Preliminary remarks

In this section first we treat the case $k=1, m=2$. The resuts concerning this case follow from the general theorems of Section 3 as well, and the reason why we deal with them separately is that many details disappear and the main line becomes clearer.

Also, we treat some simple special cases which do not follow from Section 3, and are needed to prove the general results.

Let $m=2$, $k=1$ and $f_2(n,1)=f(n)$. In other words $f(n)$ is the minimal length of a stratgegy having the property that the intersection of any two different questions is at most one. It is easily verified that $f(1)=0$, $f(2)=1$, $f(3)=f(4)=2$, $f(5)=f(6)=f(7)=3$, and clearly either $f(n+1)=f(n)$ or $f(n+1)=f(n)+1$ for all n.

Let $n_0, n_1, \ldots$ $(n_0<n_1<\cdots)$ be the 'jump-points' of the function f, i.e. $f(n_i+1)=f(n_i)+1$ and f is constant on the interval $(n_{i-1}, n_i]$ $(i=1,2,\ldots)$. (E.g. $n_0=1$, $n_1=2$, $n_2=4$, $n_3=7$.) Clearly, if $n_{i-1}<n\le n_i$ then $f(n)=i$, and thus f is given by the n_i's.

Lemma. $n_i=n_{i-1}+i$.

Proof. I. Let $|H|=n_i$ and S,

$$S(x)=(A_1, A_2(x), \ldots, A_i(x)) \quad (x\in H)$$

be an optimal strategy on H with

$$|A_{j_1}(x)\cap A_{j_2}(x)|\le 1 \quad (1\le j_1<j_2\le i).$$

(a) $(A_2(x)\cap A_1, \ldots, A_i(x)\cap A_1)$ $(x\in A_1)$ is a strategy of length $i-1$ on A_1 and all sets have at most one element.

(b) $(A_2(x)\cap(H\setminus A_1), \ldots, A_i(x)\cap(H\setminus A_1))$ $(x\in H\setminus A_1)$ is a strategy of length $i-1$ on $H\setminus A_1$, and the intersection of any two different sets is at most 1.

From (a) $|A_1|\le i$, and from (b) $|H\setminus A_1|\le n_{i-1}$ follows, thus $n_i=|H|\le n_{i-1}+i$ is proved.

II. Conversely, if a strategy of length $i-1$ is given on A_1 with sets having at most one element and a strategy of length $i-1$ on $H\setminus A_1$ satisfying (2) with $k=1$, then let A_1 be the first question, and let us take the former strategy for $x\in A_1$, and the latter one for $x\in H\setminus A_1$. So we have defined a strategy of length i on H which satisfies (2) with $k=1$. We may let $|A_1|=i$ and $|H\setminus A_1|=n_{i-1}$ which gives us a set H, $|H|=n_{i-1}+i$, with a strategy of length i, satisfying (2) with $k=1$. As n_i is the maximum cardinality of sets having such a strategy, $n_i\ge n_{i-1}+i$ and the lemma follows.

Proposition 1. $n_i=1+\frac{1}{2}i(i+1)$.

This proposition clearly follows from the lemma. The function f is determined by

Proposition 1. We want however to get an explicit formula:

$$n_i = 1 + \tfrac{1}{2}i(i+1) = \tfrac{7}{8} + \tfrac{1}{2}(i+\tfrac{1}{2})^2$$

and we know that

$$\tfrac{7}{8} + \tfrac{1}{2}(i-\tfrac{1}{2})^2 = n_{i-1} < n \le n_i = \tfrac{7}{8} + \tfrac{1}{2}(i+\tfrac{1}{2})^2$$

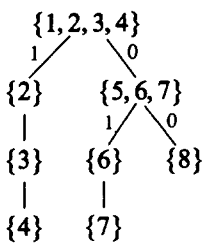
is equivalent to $f(n) = i$. So we may write $f(n)$ instead of i in the inequality and hence obtain the following upper and lower estimates for $f(n)$:

$$\sqrt{2n-\tfrac{7}{4}} - \tfrac{1}{2} \le f(n) < \sqrt{2n-\tfrac{7}{4}} + \tfrac{1}{2}.$$

This inequality yields:

Corollary. $f(n) = \lceil \sqrt{2n-\tfrac{7}{4}} - \tfrac{1}{2} \rceil$

Remark. The optimal strategy is easily constructed by virtue of the lemma (see Section 3). For example for $n=9$, $H=\{1,2,\ldots,9\}$, the following strategy is obtained:



If $k \ge n/2^m$, $f_m(n,k)$ and the optimal strategy satisfying the condition (m) can be immediately determined as the following statement shows:

Proposition 2. *If* $k \ge n/2^m$ *then* $f_m(n,k) = \lceil \log n \rceil$.

Proof. We shall construct a strategy of length $\lceil \log n \rceil$ satisfying the inequality (m) with $k = n/2^m$ for $m = 1,2,\ldots,\lceil \log n \rceil$.

We define A_i by recursion independently of $x \in H$:

Let $A_1 \subset H$, $|A_1| = \lfloor \tfrac{1}{2}n \rfloor$, and let A_2 be the union of a subset of cardinality $\lfloor \tfrac{1}{2}|H \setminus A_1| \rfloor$ of $H \setminus A_1$ and of a subset of cardinality $\lfloor \tfrac{1}{2}|A_1| \rfloor$ of A_1.

Suppose that $A_1, \ldots, A_i$ are already defined. Let $H_1^{(i)}, \ldots, H_{2^i}^{(i)}$ denote the sets of the form $A_1^\varepsilon \cap \cdots \cap A_i^\varepsilon$ ($A_j^\varepsilon = A_j$ or $A_j^\varepsilon = \bar{A}_j$). Let $B_j^{(i)} \subset H_j^{(i)}$, $|B_j^{(i)}| = \lfloor \tfrac{1}{2}|H_j^{(i)}| \rfloor$ and define A_{i+1} with $A_{i+1} = \bigcup_{j=1}^{2^i} B_j^{(i)}$. It is easy to see that $(A_1, \ldots, A_{\lceil \log n \rceil})$ is a strategy (even a static strategy) on H, and by induction on m the inequalities

$$|A_{i_1} \cap \cdots \cap A_{i_m}| \le \frac{n}{2^m} \quad (i_1 < \cdots < i_m)$$

($m=1,2,\dots,\lceil \log n \rceil$) easily follow.

In Section 3, $f_m(n,k)$ will be calculated similarly to the case $m=2$, $k=1$: For all fixed m and k either $f_m(n+1,k)=f_m(n,k)+1$ or $f_m(n+1,k)=f_m(n,k)$ and thus the function $f_m(\,,k)$ may also be given by its 'jump-places'. Let $F_m(N,k)$ be the N-th 'jump-place', i.e. $F_m(0,k)=1$, $F_m(1,k)=2$ (for all $m,k\geq 1$) and generally

$$F_m(N,k)=\max_{f_m(n,k)=N} n.$$

In Section 3, instead of $f_m(n,k)$ we will examine $F_m(N,k)$. $F_m(N,k)$ obviously determines $f_m(n,k)$, and the following inequality enables us to obtain an explicit formula for $f_m(n,k)$:

$$F_m(f_m(n,k)-1,k)<n\leq F_m(f_m(n,k),k). \qquad (*)$$

(In fact, by the definition of $F_m(N,k)$ the inequality $F_m(N-1,k)<n\leq F_m(N,k)$ is equivalent to $f_m(n,k)=N$. Writing $f_m(n,k)$ instead of N we have $(*)$.) First, let us notice that $n_i=F_2(i,1)$.

Proposition 2 has the following corollary for $F_m(N,k)$:

Corollary. *If* $N\leq m+\lfloor \log k \rfloor$ $(=\lfloor \log 2^m k \rfloor)$ *then* $F_m(N,k)=2^N$.

Proof. $N\leq \lfloor \log 2^m k \rfloor$ implies

$$2^N\leq 2^{\lfloor \log 2^m k \rfloor}\leq 2^m k,$$

and hence by Proposition 1, $f_m(2^N,k)=N$. Obviously $f_m(2^N+1,k)=N+1$ and thus the statement follows from the definition of $F_m(N,k)$.

$F_m(N,k)$ will be calculated by the recursion formulas having the numbers $F_1(N,k)$ as initial values together with the values determined in the above corollary. As mentioned in Section 1 $f_1(n,k)$ has been determined (Katona (1973)), and so the calculation of $F_1(N,k)$ does not cause any problem:

Proposition 3. *Provided* $N\geq \lfloor \log k \rfloor +1$,

$$F_1(N,k)=k(N-\lfloor \log 2k \rfloor + c_k)$$

where $c_k=2^{\lfloor \log 2k \rfloor}/k$.

Remark. 1. Obviously, $1<c_k\leq 2$ for all k.

2. For $N\leq \lfloor \log k \rfloor +1$ the equality $F_1(N,k)=2^N$ has been proved in the corollary of Proposition 2. If $N=\lfloor \log k \rfloor +1$ both formulas hold.

Proof. If $n>k$

$$f_1(n,k)=i_0+\lceil \log(n-k_{i_0}) \rceil$$

where $i_0 = \lceil n/k \rceil - 2$ (Katona (1973)). (The optimal strategy is obtained taking i_0 pairwise disjoint k-element sets $A_1, \dots, A_{i_0}$ in H and then – denoting by A the set which contains x from the sets $A_1, \dots, A_{i_0}$, $H \setminus \bigcup_{i=1}^{i_0} A_i$ – taking the trivial strategy of length $\lceil \log |A| \rceil$ on A.) The formula for $f_1(n,k)$ will be transformed for our present purposes: Let $[n]_k$ denote the number satisfying $k < [n]_k \le 2k$, $[n]_k \equiv n \pmod{k}$. Obviously

$$\left\lceil \frac{n}{k} \right\rceil - 2 = \frac{n - [n]_k}{k}.$$

Thus

$$f_1(n,k) = \left\lceil \frac{n}{k} \right\rceil - 2 + \lceil \log[n]_k \rceil = \frac{n - [n]_k}{k} + \lceil \log[n]_k \rceil$$

if $n > k$. The condition

$$N \ge \lfloor \log k \rfloor + 1 = \lfloor \log 2k \rfloor$$

implies $F_1(N,k) \ge 2^{\lfloor \log 2k \rfloor} > k$, so we see that the values n *greater than* k (for which $f_1(n+1,k) = f_1(n,k) + 1$) are to be found.

Case 1. $k \nmid n$. Then

$$\left\lceil \frac{n+1}{k} \right\rceil = \left\lceil \frac{n}{k} \right\rceil \quad \text{and} \quad [n+1]_k = [n]_k + 1.$$

Hence $f_1(n+1,k) = f_1(n,k) + 1$ if and only if $[n]_k$ is a power of 2.

Case 2. $k \mid n$. Then

$$\left\lceil \frac{n+1}{k} \right\rceil = \left\lceil \frac{n}{k} \right\rceil + 1 \quad \text{and} \quad [n]_k = 2k, \quad [n+1]_k = k+1.$$

Hence $f_1(n+1,k) = f_1(n,k) + 1$ if and only if $\lceil \log 2k \rceil = \lceil \log(k+1) \rceil$. This holds if and only if k is a power of 2.

Since $2k = [n]_k$ in Case 2, and $2k$ is a power of 2, summarizing the two cases we can state that

$$f_1(n+1,k) = f_1(n,k) + 1$$

if and only if $[n]_k$ is a power of 2. As $k < [n]_k \le 2k$ and the unique power of 2 in the interval $(k, 2k]$ is $2^{\lfloor \log 2k \rfloor}$, we have

$$f_1(n+1,k) = f_1(n,k) + 1$$

if and only if $[n]_k = 2^{\lfloor \log 2k \rfloor}$. For such n

$$f_1(n,k) = \frac{n - 2^{\lfloor \log 2k \rfloor}}{k} + \lfloor \log 2k \rfloor$$

and we have

$$n = k\left(f_1(n,k) - \lfloor \log 2k \rfloor + \frac{2^{\lfloor \log 2k \rfloor}}{k} \right).$$

This means that

$$\max_{f_1(n,k)=N} n = k\left(N - \lfloor \log 2k \rfloor + \frac{2^{\lfloor \log 2k \rfloor}}{k}\right),$$

which is our statement.

Remark: Inequality (*) together with Proposition 3 gives

$$k(f_1(n,k) - 1 - \lfloor \log 2k \rfloor + c_k) \le n \le k(f_1(n,k) - \lfloor \log 2k \rfloor + c_k)$$

hence

$$\frac{n}{k} - c_k + \lfloor \log 2k \rfloor \le f_1(n,k) < \frac{n}{k} - c_k + 1 + \lfloor \log 2k \rfloor.$$

Thus

$$f_1(n,k) = \left\lceil \frac{n}{k} - c_k \right\rceil + \lfloor \log 2k \rfloor.$$

(This must be equal to $i_0 + \lceil \log(n - i_0 k) \rceil$, $i_0 = \lceil n/k \rceil - 2$.) Now, we have all the means to treat the general results.

3. General results

The basic recursive formula is given in the following lemma:

Lemma. *If $m \ge 2$ and $N, k \ge 1$, then*

$$F_m(N,k) = F_m(N-1,k) + F_{m-1}(N-1,k).$$

Proof. I. Let $|H| = F_m(N,k)$ and S,

$$S(x) = (A_1, A_2(x), \ldots, A_N(x)) \quad (x \in H),$$

be an optimal strategy on H satisfying (m).

(a) S_{A_1}:

$$S_{A_1}(x) = (A_2(x) \cap A_1, \ldots, A_N(x) \cap A_1) \quad (x \in A_1)$$

is clearly a strategy of length $N-1$ on A_1 and satisfies $(m-1)$.

(b) $S_{H \setminus A_1}$:

$$S_{H \setminus A_1}(x) = (A_2(x) \cap (H \setminus A_1), \ldots, A_N(x) \cap (H \setminus A_1)) \quad (x \in H \setminus A_1)$$

is obviously a strategy of length $N-1$ on $H \setminus A_1$ and satisfies (m).

Thus

$$|A_1| \le F_{m-1}(N-1,k), \qquad |H \setminus A_1| \le F_m(N-1,k).$$

II. Conversely, if A_1 is a subset of H and S_{A_1}:

$$S_{A_1}(x) = (A'_1(x), A'_2(x), \ldots, A'_{N-1}(x)) \quad (x \in A_1)$$

is a strategy of length $N-1$ on A_1 satisfying $m-1$ and $S_{H\setminus A_1}$:

$$S_{H\setminus A_1}(x)=(A_1''(x),\dots,A_{N-1}''(x)) \quad (x\in H\setminus A_1)$$

is a strategy of length $N-1$ on $H\setminus A_1$ satisfying (m), then S:

$$S(x)=(A_1,A_2(x),\dots,A_N(x))$$

defined with

$$A_i(x)=A_{i-1}'(x) \quad \text{if } x\in A_1,$$

$$A_i(x)=A_{i-1}''(x) \quad \text{if } x\in H\setminus A_1$$

$(2\le i\le N)$, is obviously a strategy on H, and satisfies (m).

Thus if H is maximal then

$$|A_1|=F_{m-1}(N-1,k), \qquad (H\setminus A_1)=F_m(N-1,k)$$

and so

$$F_m(N,k)=|H|=F_m(N-1,k)+F_{m-1}(N-1,k)$$

and the theorem is proved.

Corollary. *If $m\ge 2$ and $N,k\ge 1$, then*

(a) $$F_m(N,k)=1+\sum_{i=0}^{N-1}F_{m-1}(i,k)$$

and

(b) $$F_m(N,k)=F_1(N-m+1,k)+\sum_{i=1}^{m-1}F_{i+1}(N-m+i,k),$$

provided $N\ge m$.

Both (a) and (b) immediately follow from the lemma.

Remark. Once $F_m(N,k)$ is known (m,k are fixed), the construction of the optimal strategy does not cause any problem with the help of the above lemma:

If $F_m(N-1,k)<n\le F_m(N,k)$, let $A_1\subset H$, $|A_1|=F_{m-1}(N-1,k)$. Then

$$|H\setminus A_1|\le F_m(N,k)-F_{m-1}(N-1,k)=F_m(N-1,k)$$

follows from the lemma. In other words it is possible to construct a strategy of length $N-1$ on A_1 satisfying $(m-1)$, and a strategy of length $N-1$ on $H\setminus A_1$ satisfying (m). By the proof of the lemma these two strategies yield a strategy of length N on H satisfying (m).

We continue in this way: For $x\in A_1$ let

$$A_2(x)\subset A_1, \quad |A_2(x)|=F_{m-1}(N-2,k)$$

and for $x\in H\setminus A_1$

$$A_2(x)\subset H\setminus A, \quad |A_2(x)|=F_m(N-2,k)$$

etc., we always have to give the 'first' question-set under the 'new' conditions. If $m=1$ or $N \le m + \lfloor \log k \rfloor$ the strategy is given in Proposition 2 and 3 respectively.

(The precise formulation of the construction can be given by induction on N.)

Unfortunately we could not get an exact explicit formula for $F_m(N,k)$ if $m \ge 3$. In this case using part (b) of the above corollary, $m^2\sqrt{n}$ additions – the result of all of which is to be compared with n and then must be stored – are enough to determine $f_m(n,k)$ for some fixed m, n, k, and to have all values $F_m(N,k)$ which are necessary for the construction of the optimal strategy. (We have got this result with the help of part (b) of the corollary to Theorem 2, and examining the algorithm in details.)

In the proof of the explicit formulas and bounds proved in the remaining part of the paper only part (a) of the corollary will be used.

Recall that $c_k = 2^{\lfloor \log 2k \rfloor}/k$ and $1 < c_k \le 2$.

Theorem 1. *Provided* $N \ge \lfloor \log 4k \rfloor$ $(= \lfloor \log k \rfloor + 2)$,

$$F_2(N,k) = \tfrac{1}{2}k((N - \lfloor \log 4k \rfloor + b_k)^2 + a_k)$$

where $b_k = c_k + \frac{1}{2}$ *and* $a_k = -c_k^2 + 3c_k - \frac{1}{4}$.

Remarks. 1. From $1 < c_k \le 2$ we get $\frac{7}{4} \le a_k \le 2$ and $\frac{3}{2} < b_k \le \frac{5}{2}$ for all k

2. For $N \le \lfloor \log k \rfloor + 2$, $F_2(N,k)$ is determined by the corollary to Proposition 2. For $N = \lfloor \log k \rfloor + 2$ both formulas hold.

Proof. By part (a) of the corollary of the lemma we have

$$F_2(N,k) = 1 + \sum_{i=0}^{N-1} F_1(i,k) = 1 + \sum_{i=0}^{\lfloor \log 2k \rfloor - 1} F_1(i,k) + \sum_{i=\lfloor \log 2k \rfloor}^{N-1} F_1(i,k)$$

$$= 2^{\lfloor \log 2k \rfloor} + \sum_{i=\lfloor \log 2k \rfloor}^{N-1} k(i - \lfloor \log 2k \rfloor + c_k).$$

(The value of $F_1(i,k)$ is given by the corollary of Proposition 2 for $i \le \lfloor \log k \rfloor$ $(= \lfloor \log 2k \rfloor - 1)$ and by Proposition 3 for $i \ge \lfloor \log 2k \rfloor$.)

$$\sum_{i=\lfloor \log 2k \rfloor}^{N-1} k(i - \lfloor \log 2k \rfloor + c_k)$$
$$= k\left(\sum_{i=1}^{N-1} i - \sum_{i=1}^{\lfloor \log 2k \rfloor - 1} i - (N - \lfloor \log 2k \rfloor)(\lfloor \log 2k \rfloor - c_k)\right)$$

and hence

$$F_2(N,k) = 2^{\lfloor \log 2k \rfloor} + k(\tfrac{1}{2}N(N-1) - \tfrac{1}{2}\lfloor \log 2k \rfloor(\lfloor \log k \rfloor - 1)$$
$$- (N - \lfloor \log 2k \rfloor)(\lfloor \log 2k \rfloor - c_k))$$
$$= \tfrac{1}{2}k((N - \lfloor \log 4k \rfloor + c_k + \tfrac{1}{2})^2 - c_k^2 + 3c_k - \tfrac{1}{4}),$$

which was to be proved.

We have arrived at our main result:

Corollary. *If* $k<\frac{1}{4}n$ *then*

$$f_2(n,k)=\lceil\sqrt{(2n/k)-a_k}-b_k\rceil+\lfloor\log 4k\rfloor$$

where a_k *and* b_k *are the numbers defined in the theorem.*

Remark. As $\frac{7}{4}\le a_k\le 2$ and $\frac{3}{2}<b_k\le\frac{5}{2}$ we may deduce that

$$f_2(n,k)=\lceil\sqrt{2n/k}\rceil+\lfloor\log k\rfloor-1+\varepsilon\quad(\varepsilon=\pm 1 \text{ or } 0).$$

Proof. From $n>4k\ge 2^{\lfloor\log 4k\rfloor}$ follows

$$f_2(n,k)\ge\lceil\log n\rceil\ge\lfloor\log 4k\rfloor+1.$$

Thus Theorem 1 applies for $N=f_2(n,k)-1$ and $N=f_2(n,k)$.

Hence, by (*) of Section 2

$$\tfrac{1}{2}k((f_2(n,k)-1-\lfloor\log 4k\rfloor+b_k)^2+a_k)$$
$$<n\le\tfrac{1}{2}k((f_2(n,k)-\lfloor\log 4k\rfloor+b_k)^2+a_k),$$

thus

$$\sqrt{(2n/k)-a_k}-b_k+\lfloor\log 4k\rfloor$$
$$\le f_2(n,k)<\sqrt{(2n/k)-a_k}-b_k+1+\lfloor\log 4k\rfloor$$

as our corollary states.

We are going to estimate now $F_m(N,k)$ in order to estimate $f_m(n,k)$.

Theorem 2. *For all* $N,k,m\in\mathbb{Z}$

$$\frac{k}{m!}(N-\lfloor\log k\rfloor-m+1)^m<F_m(N,k)\le\frac{k}{m!}(N-\lfloor\log k\rfloor+2)^m$$

provided $N\ge m+\lfloor\log k\rfloor$.

Proof. We proceed by induction on m: For $m=1$ the statement immediately follows from Proposition 3 (for $m=2$ it follows from Theorem 1).

Suppose the theorem is true for m and then let us prove it for $m+1$. To do this we will need the inequality

$$\frac{r^{t+1}}{t+1}\le\sum_{i=1}^{r}i^t\le\frac{(r+1)^{t+1}}{t+1}$$

which follows from the estimation of a sum with an integral ($t\ge 0$ is a real number and $r\in\mathbb{Z}$).

By part (a) of the corollary to the lemma we have

$$F_{m+1}(N,k)=1+\sum_{i=0}^{N-1}F_m(i,k)=1+\sum_{i=0}^{m+\lfloor\log k\rfloor-1}F_m(i,k)+\sum_{i=m+\lfloor\log k\rfloor}^{N-1}F_m(i,k)$$

$$=2^{m+\lfloor\log k\rfloor}+\sum_{i=m+\lfloor\log k\rfloor}^{N-1}F_m(i,k). \qquad (\bigstar)$$

I. Lower estimate. By ($\bigstar$)

$$F_{m+1}(N,k)>\sum_{i=m+\lfloor\log k\rfloor}^{N-1}F_m(i,k).$$

Thus, using the lower estimate for m,

$$F_{m+1}(N,k)>\frac{k}{m!}\sum_{i=m+\lfloor\log k\rfloor}^{N-1}(i-\lfloor\log k\rfloor-m+1)^m$$

$$=\frac{k}{m!}\sum_{j=1}^{N-\lfloor\log k\rfloor-m}j^m\geq\frac{k}{m!}\,\frac{(N-\lfloor\log k\rfloor-m)^{m+1}}{m+1}$$

and this is the lower estimate for the case $m+1$. (In the last step we have used the inequality $\sum_{i=1}^r i^t\geq r^{t+1}/(t+1)$.)

II. Upper estimate

$$2^{m+\lfloor\log k\rfloor}=2^{\lfloor\log 2^m k\rfloor}\leq 2^m k.$$

Thus, by ($\bigstar$), using the upper estimate for m:

$$F_{m+1}(N,k)\leq 2^m k+\frac{k}{m!}\sum_{i=m+\lfloor\log k\rfloor}^{N-1}(i-\lfloor\log k\rfloor+2)^m$$

$$=2^m k-\frac{k}{m!}\sum_{j=1}^{m+1}j^m+\frac{k}{m!}\sum_{j=1}^{N-\lfloor\log k\rfloor+1}j^m$$

$$\leq k\left(2^m-\frac{(m+1)^{m+1}}{(m+1)!}+\frac{k}{(m+1)!}\,(N-\lfloor\log k\rfloor+2)^{m+1}\right).$$

(In the last step we used the inequality $\sum_{i=1}^r i^t\leq(r+1)^{t+1}/(t+1)$.)

The upper estimate is then proved for $m+1$ since

$$2^m\leq\frac{(m+1)^{m+1}}{(m+1)!}\quad(m\in\mathbb{Z}).$$

This follows by induction on m: for $m=1$ equality holds; denoting

$$p_m=\frac{(m+1)^{m+1}}{(m+1)!}$$

we have

$$p_m=\left(1+\frac{1}{m}\right)^{m+1}p_{m-1}>\left(1+\frac{1}{m}\right)^m p_{m-1}\geq 2p_{m-1}\quad(m\geq 2)$$

which proves the inequality.

Corollary. *For all $N, k, m \in \mathbb{Z}$:*

$$\left(m!\,\frac{n}{k}\right)^{1/m} + \lfloor \log k \rfloor - 2 \le f_m(n,k) < \left(m!\,\frac{n}{k}\right)^{1/m} + \lfloor \log k \rfloor + m$$

provided $k < n/2^m$.

Remark. The above interval contains $m+2$ integers. For $m=2$, $f_m(n,k)$ may fall near both the upper and lower estimates as the remark after the corollary of Theorem 1 shows. This is probably true for every m.

Proof. From $n > 2^m k \ge 2^{m+\lfloor \log k \rfloor}$ it follows that

$$f_m(n,k) \ge \lceil \log n \rceil \ge m + \lfloor \log k \rfloor + 1,$$

thus Theorem 2 applies for $N = f_m(n,k)$. Hence by (*) of Section 2

$$n \le F_m(f_m(n,k),k) \le \frac{k}{m!}\,(f_m(n,k) - \lfloor \log k \rfloor + 2)^m$$

and

$$n > F_m(f_m(n,k)-1,k) < \frac{k}{m!}\,(f_m(n,k) - \lfloor \log k \rfloor - m)^m$$

and the statement immediately follows.

The inequalities used in the proof of Theorem 2 are quite rough but using finer inequalities the estimates for $F_m(N,k)$ will not become considerably sharper. However, the recursion formulas of Section 3 enable us to calculate the exact value of $F_m(N,k)$ and thus $f_m(n,k)$ for all fixed m, N, k and m, n, k respectively and make clear the construction of optimal strategies.

Acknowledgement

The author wishes to thank Professor G.O.H. Katona for his encouragement and for the very careful reading of the paper.

References

Rényi, A. (1965). On the theory of random search, *Bull. Am. Math. Soc.* 71, 809–828.

Katona, G.O.H. (1966). On separating systems of finite sets, *J. Comb. Theory* 1, 174–194.

Katona, G.O.H. (1973). Combinatorial search problems. In: J.N. Srivastava, Ed., *A Survey on Combinatorial Theory*. North-Holland, Amsterdam, 296–297.

Katona, G.O.H. (1976). Search using sets with small intersection, *Colloques Internat. C.N.R.S.* No. 276, Theorie de l'Information.