Indexed
fay

A7077
A7078

INDIANA UNIVERSITY
AT SOUTH BEND

1700 MISHAWAKA AVENUE
P.O. BOX 7111
SOUTH BEND, INDIANA 46634

DIVISION OF
LIBERAL ARTS AND SCIENCES
Northside Hall
TEL: (219) 237-4214
FAX: (219) 237-4538

July 25, 1991

Neil Sloane
ATT Bell Labs, Room 2C-376
600 Mountain Avenue
Murray Hill, NJ 07974

Dear Mr. Sloane:

I am responding to your letter of June 23, 1991, a copy of which is enclosed. The work I have been doing involves minimizing the cost of search in ordered arrays with variable probe costs. Almost all the sequences that have turned up in this work are *derived from* recursively defined, two-variable sequences having one or the other of the two forms shown below. In each formula, $P(k)$ is a strictly positive, strictly increasing "penalty" function defined on positive integers k. The variables n and t are non-negative integers.

(1) $$S(n,t) = \min_{1 \le r \le n} \{ P(r+t)\,n + S(r-1,t) + S(n-r, r+t) \},$$
$$S(0,t) = 0 \quad \text{for all } t \ge 0.$$

(2) $$M(n,t) = \min_{1 \le r \le n} \{ P(r+t) + \max\{ M(r-1,t),\ M(n-r, r+t) \} \},$$
$$M(0,t) = 0 \quad \text{for all } t \ge 0.$$

What I'm actually looking for is formulas for the one-variable sequences $S(n,0)$ and $M(n,0)$ for $n \ge 1$. Examples of these sequences for various choices of $P(k)$ are given on the next page.

The way these sequences arise is that we suppose we are searching an array of length n whose entries are "no" and "yes". If an array entry is "yes", then all entries to its right are "yes" as well, so in general the array consists of a string of "no"s followed by a string of "yes"s. The problem is to find the location of the first "yes" (if there is one). We also assume that probing the k-th location in the array requires an amount of time given by $P(k)$. This arises in a problem in filter design, where it is necessary to estimate how many variables will be needed to solve a certain linear programming problem, and where it is desirable to have as few variables as possible. This means trying out different numbers of variables and geting "no" or "yes" answers; "trying a number of variables" means trying to solve a linear programming problem using only that many variables, which typically takes an amount of time proportional to the number of variables.

A7077

A7078

√ f_{94}

$P(k) = k$

n	S(n,0)	M(n,0)
1	1	1
2	4	3
3	10	5
4	19	7
5	31	9
6	47	12
7	68	15
8	92	19
9	120	23
10	153	26
11	190	29
12	232	32
13	279	35
14	332	38
15	392	41
16	454	45
17	521	49
18	593	53
19	670	57
20	753	62

I regard this as the most interesting of the sequences.

$P(k) = 2^k$

n	S(n,0)	M(n,0)
1	2	2
2	8	6
3	22	12
4	50	24
5	110	48
6	226	96
7	464	192
8	938	384
9	1888	768
10	3794	1536
11	7598	3072
12	15208	6144
13	30438	12288
14	60890	24576
15	121792	49152
16	243606	98304
17	487238	196608
18	974488	393216
19	1948998	786432
20	3898034	1572864

Trivially $M(n,0) = 3 \cdot 2^{n-1}$ for $n > 1$.

$P(k) = k!$

n	S(n,0)	M(n,0)
1	1	1
2	4	3
3	13	8
4	45	30
5	197	144
6	1069	840
7	6981	5760
8	53207	45360
9	462313	403200
10	4500208	3991680
11	48454894	43545600
12	5.714E08	5.189E08
13	7.321E09	6.706E09
14	1.012E11	9.341E10
15	1.503E12	1.395E12
16	2.383E13	2.223E13
17	4.018E14	3.766E14
18	7.182E15	6.758E15
19	1.356E17	1.280E17
20	2.697E18	2.555E18

(1.234E05 = 1.234×10^5)

Trivially $M(n,0) = n! + (n-1)!$ for $n > 1$.

%N Optimal cost of search tree.
%R SIAC 17 1213 88.
%O 1,2

for both

The problems I have worked on all involve selecting a penalty function and then trying to find an "optimal" search strategy for finding the first "yes" in the array. There are two ways to decide whether one strategy is better than another: the first compares the *expected* amounts of time required by the two strategies; the second compares the *maximum* amounts of time required. In the notation used above, S (n, 0) is the expected amount of time required by an optimal strategy in the expected value sense to find the first "yes" in an array of length n . Similarly, M (n, 0) is the maximum amount of time required by an optimal strategy in the "minimax" sense. All this is explained in more detail in the reprint and preprint I have enclosed.

The reprint shows that when $P(k) = k$, the sequence S (n, 0) is asymptotic with $\frac{1}{2}(n + 1)^2 \lg (n + 1)$, but I was not able to get an asymptotic formula for the much tamer looking sequence M (n, 0) . A colleague and I have worked on this quite a bit, and we cannot "capture" that sequence. ***If the sequence is familiar to you, or if you can see what its asymptotic behavior is, we would very much like to hear from you!***

The preprint I have enclosed shows that when $P(k) = 2^k$, the sequence S (n, 0) is $\Theta(2^k)$, where as usual, $f(n) = \Theta(g(n))$ iff there exist positive constants A and B such that for all large n , $A|g(n)| \le |f(n)| \le B|g(n)|$. Also, we have an *exact* formula for M (n, 0) in this case:

$$(*) \qquad M(n, 0) = P(n) + P(n-1) \text{ for all } n > 1, \qquad M(1, 0) = P(1).$$

In fact, formula (*) is valid for every penalty function P(k) that satisfies the inequality

$$P(k) \ge P(k-2) + P(k-3) \quad \text{for all } k \ge 3.$$

This includes $P(k) = k!$ and all penalty functions of the form $P(k) = b^k$ in which the constant b exceeds 1.325 (the approximate root of $b^3 = b + 1$) . Finally, when $P(k) = k!$, the sequence S (n, 0) is asymptotic to $n!$, or, if more precision is desired, to $n! + 2(n-1)! + 3(n-2)! + 4(n-3)!$.

As I indicated, the most interesting and puzzling of these sequences is M (n, 0) when $P(k) = k$. Computer calculations suggest that for all large n it is bounded above by $C n \lg n$ for some constant C in the vicinity of 0.8 , but we have not been able to prove this, nor have we found a good lower bound formula for M (n, 0) .

I hope you find one or more of these sequences sufficiently interesting to include them in your book. If I can answer any further questions, I will be happy to do so.

Sincerely,

Wm J. Knight

William J. Knight

Scan 7077
Knight 7078
letter

3 ~~R~~ pages
2 seqs