

Exercise 3, Discrete Mathematics for Bioinformatics

Sascha Meiers, Martin Seeger

Winter term 2011/2012

3.1 Skip lists

a) Expected value of h (adapted from script): we use the notation from the script: $x \in S$, $h(x)$ = number of sets S_i containing x , $h = 1 + \max\{h(x) : x \in S\}$.

For $k \geq 1$, we have $P(h(x) \geq k) = p^{k-1}$ and therefore

$$P(h \geq k+1) = nP(h(x) \geq k) = np^{k-1}.$$

This estimate does not make sense for $k < 1 + \log_{1/p} n = 1 - \log_p n$. For those values of k we can use the trivial upper bound $P(h \geq k+1) \leq 1$. Then $E(h)$ equals:

$$\begin{aligned} \sum_{k=1}^{\infty} P(h \geq k+1) &= \sum_{k=1}^{\lceil -\log_p n \rceil} P(h \geq k+1) + \sum_{k=1+\lceil -\log_p n \rceil}^{\infty} P(h \geq k+1) \leq \\ &\leq 1 + \lceil -\log_p n \rceil + \sum_{k=1+\lceil -\log_p n \rceil}^{\infty} np^{k-1} = \\ &= 1 + \lceil -\log_p n \rceil + \frac{np^{\lceil -\log_p n \rceil}}{1-p} = \\ &\leq 1 + \lceil -\log_p n \rceil + \frac{np^{-\log_p n}}{1-p} = \\ &= 1 + \lceil -\log_p n \rceil + \frac{1}{1-p}. \end{aligned}$$

For $p = 1/3$ this yields $E(h) \leq 5/2 + \lceil \log_3 n \rceil$.

b) Expected value of space consumption (adapted from script): let M denote the total size of the sets S_1, S_2, \dots, S_h . Then $M = \sum_{x \in S} h(x)$ and by linearity of expectation:

$$E(M) = \sum_{x \in S} E(h(x)) = \frac{n}{p}.$$

We need to add the h pseudo nodes at $-\infty$, so that the total size is

$$E(M) + E(h) \leq \frac{n}{p} + 1 + \lceil -\log_p n \rceil + \frac{1}{1-p}.$$

For $p = 1/3$ this yields $E(M) + E(h) \leq 3n + 5/2 + \lceil \log_3 n \rceil$.

c) Expected value of search time (adapted from script): Let x be a real number and let C_i denote the number of elements in the list L_i that are inspected when searching for x . (We do

not count the element of L_i at which the algorithm starts walking to the right. Hence, C_i counts comparisons between x and elements of S .) The search cost is then proportional to $\sum_{i=1}^h (1 + C_i)$.

We first estimate the search level above A , i.e., the total costs in the lists $L_{A+1}, L_{A+2}, \dots, L_h$. Since the cost is at most equal to the total size of these lists, its expected value is at most equal to the expected value of $M_A := \sum_{i=A+1}^h |L_i|$.

We can write

$$E(M_A) = \sum_{k=0}^n E(M_A | |S_{A+1}| = k) P(|S_{A+1}| = k),$$

where

$$E(M_A | |S_{A+1}| = k) = 2k,$$

and

$$P(|S_{A+1}| = k) = \binom{n}{k} p^{Ak} (1 - p^A)^{n-k}.$$

Therefore

$$E(M_A) = 2 \sum_{k=0}^n k \binom{n}{k} p^{Ak} (1 - p^A)^{n-k} = 2np^A,$$

for lists above A .

For lists up to A , we consider

$$E(C_i) = \sum_{k=1}^n E(C_i | l_i(x) = k) P(l_i(x) = k),$$

where $l_i(x)$ is the number of elements in L_i that are $\leq x$.

We have for the first term

$$E(C_i | l_i(x) = k) = \sum_j j P(C_i = j | l_i(x) = k) \leq \sum_j j p^{j-1} = (1 - p)^{-2}$$

independently of k whence follows that

$$E(C_i) \leq (1 - p)^{-2} \sum_{k=1}^n P(l_i(x) = k) = (1 - p)^{-2}.$$

For the search cost up to A we obtain

$$E \left(\sum_{i=1}^A (1 + C_i) \right) \leq A(1 + (1 - p)^{-2}).$$

Adding up, this yields total expected cost

$$E \left(\sum_{i=1}^A (1 + C_i) \right) + E(M_A) \leq A(1 + (1 - p)^{-2}) + 2np^A.$$

Using $A = \log_{1/p} n$, this becomes

$$E \left(\sum_{i=1}^A (1 + C_i) \right) + E(M_A) \leq (1 + (1 - p)^{-2}) \log_{1/p} n + 2,$$

and setting $p = 1/3$, we find

$$E \left(\sum_{i=1}^A (1 + C_i) \right) + E(M_A) \leq \frac{13}{4} \log_3 n + 2.$$

3.2 “Sparse” skip list

a) x

3.3 Skip lists

a) x

3.4 Independencies

We have

$$E(X_1) = \frac{1}{9}(1 + 1 + 2 + 2 + 3 + 3 + 1 + 2 + 3) = 2,$$

$$E(X_2) = \frac{1}{9}(2 + 3 + 1 + 3 + 1 + 2 + 1 + 2 + 3) = 2,$$

$$E(X_3) = \frac{1}{9}(3 + 2 + 3 + 1 + 2 + 1 + 1 + 2 + 3) = 2.$$

i) x

ii) x

iii) x

iv) x

v) x

vi) $N = X_2$, $E(N) = 2$. Therefore,

$$\sum_{i=1}^{E(N)} E(X_i) = E(X_1) + E(X_2) = 4.$$

On the other hand,

$$\begin{aligned} E\left(\sum_{i=1}^N X_i\right) &= P(N=1)E\left(\sum_{i=1}^1 X_i \middle| N=1\right) + P(N=2)E\left(\sum_{i=1}^2 X_i \middle| N=2\right) + \\ &+ P(N=3)E\left(\sum_{i=1}^3 X_i \middle| N=3\right) = \frac{2}{3} + \frac{2+2}{3} + \frac{2+2+3}{3} = \frac{13}{3}. \end{aligned}$$