Exercise 3, Discrete Mathematics for Bioinformatics

Sascha Meiers, Martin Seeger

Winter term 2011/2012

3.1 Skip lists

a) Expected value of h: 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:

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

b) Expected value of search time:

c) Expected value of space consumption:

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,

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