Assignment 2

Advanced Algorithms and Datastructures

Authors:
Jenny-Margrethe Vej (rwj935)
Martin Gielsgaard Grünbaum (wrk272)
Martin Nicklas Jørgensen (tzk173)

May 22, 2014

1 Hash functions for sampling

1.1 Exercise 1(a')

We must show that $p \leq Pr[h_m(x)/m < p] \leq 1.01p$.
We are given some $p \geq 100/m$, a suitably large m and a strongly universal hashing function $h_m : U \rightarrow [m]$. In the following, we make use of the fact that $p \geq 100/m$ implies that $p/100 \geq 1/m$ and that $a \leq \lceil a \rceil < a + 1$.
We observe that

$$\begin{aligned}
& Pr[h_m(x)/m < p] && (1)\\
&= \sum_{0 \leq k < pm} Pr[h_m(x) = k] && (2)\\
&= \sum_{0 \leq k < pm} \frac{1}{m} && (3)\\
&= \frac{1}{m} \sum_{0 \leq k < pm} 1 && (4)\\
&= \frac{1}{m} \left|[0, pm)\right| && (5)\\
&= \frac{1}{m} \cdot \lceil pm \rceil && (6)\\
&= \frac{\lceil pm \rceil}{m} && (7)
\end{aligned}$$

1.2 Exercise 1(b)

2 Bottom-k sampling

2.1 Exercise 2

2.2 Exercise 3(a)

We would store the buttom-k samples in a minimum heap structure H, sorted by their hashing value. This way we can insert new entries in $O(\lg\ n)$, and retrieve the $S_h^k(H)$ lowest hash values in $O(k\ \lg\ n)$ where n is the total number of input values.

2.3 Exercise 3(b)

As written above we would be able to process/insert the next key in $O(\lg\ n)$ time.

2.4 Exercise 4

2.5 Exercise 4(a)

We will prove the equality $S_h^k(\mathrm{A} \cup \mathrm{B}) = S_h^k(S_h^k(A) \cup S_h^k(B))$. We can see each set as a sorted stack that keeps the smallest values at the top. The left hand part of the equality ($S_h^k(\mathrm{A}\cup\mathrm{B})$)) corresponds

to merging the two stacks and taking the k top values. The right hand side $(S_h^k(S_h^k(A) \cup S_h^k(B)))$ corresponds to taking the k topmost values from both stacks and then merging them and taking the k smallest values from the resulting stack.
Since we take the k smallest values from each stack we are guaranteed to have the smallest value from the union of A and B since even if the smallest values in B is bigger than the biggest values in A we still have the k smallest values from A, thus proving the equality.

2.6 Exercise 4(b)

2.7 Exercise 4(c)

3 Bottom-k sampling with strong universality

3.1 Exercise 5

3.2 Exercise 6

3.3 Exercise 7