

# Assignment 5 \*

Ishaan Kapoor

February 25, 2022

## 1 Streaming Algorithms

### A

S1:

Labels: [g,l,r,w,p,s]

Counter:[1,3,1,250379,1,577089]

Count/m: [0,0,0,0.062595,0,0.14427225]

S2:

Labels:[h,e,a,p,g,o]

Counter:[870770,1,457613,1,266870,1]

Count/m: [0.174154, 0, 0.01915226, 0, 0.053374, 0]

We know,

$$f_j - \epsilon m \leq \hat{f}_j \leq f_j$$

here,

$$\epsilon = \frac{1}{k}$$

$$\therefore f_j \leq \hat{f}_j + \epsilon m$$

For S1,  $k = 7$ ,  $m = 4000000$

Corrected  $f_j = [14.28\%, 14.28\%, 14.28\%, 20.54\%, 14.28\%, 28.71\%]$

$\therefore$  no object occurs at least 40% or even 30% of the time as the above list gives us the higher estimate

Similarly, For S2

Corrected  $f_j = [31.70\%, 14.28\%, 23.43\%, 14.28\%, 19.62\%, 14.28\%]$

Here, 'h' must occur for more than 30% of times (because h was part of the starting label list as well) however no other term can even possibly occur more than 30% of times

### B

For S1

a=52691

b=124105

c=124105

$$\begin{aligned}(\text{Count}/m)_a &= 0.013 \\ (\text{Count}/m)_b &= 0.031 \\ (\text{Count}/m)_c &= 0.031\end{aligned}$$

For S2

$$\begin{aligned}a &= 944005 \\ b &= 155553 \\ c &= 195767\end{aligned}$$

$$\begin{aligned}(\text{Count}/m)_a &= 0.18 \\ (\text{Count}/m)_b &= 0.031 \\ (\text{Count}/m)_c &= 0.039\end{aligned}$$

$$\begin{aligned}\text{We Know,} \\ P[X > 20\%] &> \frac{E(X)}{\alpha} \\ \therefore \text{for } X \geq 20\% \dots\end{aligned}$$

We know,

$$\begin{aligned}k &= \frac{2}{\epsilon} \\ \implies \epsilon &= \frac{2}{9}\end{aligned}$$

Also,

$$\hat{f}_j - \epsilon m \geq f_j$$

For S1,

$$\begin{aligned}-836197 &\leq f_a \leq 52691 \\ -764783 &\leq f_b \leq 124105 \\ -764783 &\leq f_c \leq 124105\end{aligned}$$

No value can be less than 20%

For S2,

$$\begin{aligned}-167106 &\leq f_a \leq 944005 \\ -955558 &\leq f_b \leq 155553 \\ -915344 &\leq f_c \leq 195767\end{aligned}$$

No value can be less than 20%

## C

Implementing Misra Gires in this way is straight forward, just use a word as a label instead of the character. However, to implement Count Min Sketch, we will need to create hash functions that takes in a word and returns a hash value between 0 to t-1. Apart from that, the implementation will be the same.

## D

While in Misra Gries we only maintain the count of some labels, and the rest are returned as 0, in Count-Min-Sketch, we usually get a value of count for almost all the members.