

Assignment 5

Algorithms for Big Data

Lavinia Pulcinella
Student ID: i6233926

March 2020

1 Graded Exercise 5.2

The main idea to select a random element in the stream with probability $\frac{1}{m}$ is to “keep track” of the current selected element.

Hence, we initially store the first element of the stream as our pick. As we move onto the next element of the stream, we select it with probability $\frac{1}{m}$.

However, m i.e. the length of the stream is usually unknown.

On the other hand, a probability of selection $\frac{1}{m}$ implies that all elements of the stream should have the same probability of selection i.e. the i -th element is sampled uniformly at random (in a uniform distribution every element has the same chance of being sampled). Thus, at every step we will use the the length of the stream up to the i -th element leading to a probability of sampling the i -th token of $\frac{1}{i}$.

Algorithm 1 Select random element in a stream

INIT

Pick = a_1

PROCESS STREAM

Choose a random element from 1 to i with probability $\frac{1}{i}$

if element is selected **then**

 Pick = a_i

Otherwise we do nothing (with probability $\frac{1}{i-1}$)

end

RETURN

Return Pick

We can prove (by induction) that the probability of selecting the i -th token is actually $\frac{1}{i}$.

Hence we want to prove:

$$P[\text{selected} = a_i] = \frac{1}{i}$$

Assuming this holds for $i - 1$. At the i -th step for any $j \leq i - 1$ we have

$$\begin{aligned}
 P[\text{pick} = a_j \text{ at } i\text{-th step}] &= P[\text{pick} = a_j \text{ **after** } i\text{-th step}] * P[a_j \text{ not chosen at } i\text{-th step}] = \\
 &= \frac{1}{i-1} \left(1 - \frac{1}{i}\right) = \\
 &= \frac{1}{i-1} \left(\frac{i-1}{i}\right) = \\
 &= \frac{1}{i}
 \end{aligned}$$

Thus, the amount of space used by the algorithm depends on the tokens of the stream which are from the (known) universe $[n] := 1, 2, \dots, 3$. Hence, since we only store one variable, the algorithm uses only a constant number of space which is $O(\log n)$.