# QUIZ-12 SOLUTION

1) → Store all the first s element of the stream to S.

→ Suppose we have seen n-1 elements, now the $n^{th}$ element arrives ($n > s$)

- with probability $\frac{s}{n}$, keep the $n^{th}$ element else discard it.
- if we picked the $n^{th}$ element, then it replaces one of the s elements in the sample S, picked uniformly at random.

Assumption :- [0.5]

- We assume that ~~positions~~ after n elements, the sample contains each element seen so far with probability $\frac{s}{n}$

- Our goal :- We need to show that after seeing element $n+1$ the sample maintains the property i.e the sample contains each element seen so far with probability $\frac{s}{n+1}$

[0.5] - Base Case - ~~when~~ when $n = s$ elements, the sample S has the desired property ⇒ Probability = $\frac{s}{n+1}$

[0.5] - Inductive Hypothesis - After n elements, the sample S contains each element seen so far with probability : $\frac{s}{n}$

[0.5] - Inductive step -

Now element $n+1$ arrives,

for elements already in S, probability that algorithm keeps it in s is :-

$$\left(1 - \frac{s}{n+1}\right) + \left(\frac{s}{n+1}\right) \cdot \left(\frac{s-1}{s}\right) = \frac{n}{n+1}$$

Example :- [1 POINT]

stream: a, b, c, d, e, f, g ... and sample size 2 :-

at the very begining, b is in the sample since $\frac{2}{2} = 1$

when time $n = 3$,

the probability of putting c into the sample is $\frac{2}{3}$,

So that the probability of keeping a is

$$P(\text{do not add c}) + P(\text{b is removed} \mid \text{add c}) \times P(\text{add c})$$

$$= \frac{1}{3} + \frac{1}{2} \times \frac{2}{3} = \frac{2}{3} = \frac{s}{n+1}$$

2) Suppose we have y darts and x target [1.5 POINTS].

- The prob. of a specific dart cannot hit the specific target is $\frac{x-1}{x}$

- y darts all fail to hit a specific target $\left(\frac{x-1}{x}\right)^y$

$$= \left(1 - \frac{1}{x}\right)^{x\frac{y}{x}} \simeq e^{-y/x}$$

Now,

- suppose we have n bits in array m elements in the set S, k hash function

- we have $x = n$ targets and $y = k \times m$ darts.

- ∴ the probability that a bit is still not hit by any dart is $e^{-\frac{km}{n}}$

- false positive rate $= \left(1 - e^{-\frac{km}{n}}\right)^k$

To get optimal k, we take the derivate of $f(k) = \left(1 - e^{-\frac{km}{n}}\right)^k$

we can calculate optimal value of k when f(k) reaches the minimal value

$$k = \frac{n}{m} \ln 2 \longrightarrow \text{[1.5 POINTS]}$$

3) – The probability (p) that some elements have at least r trailing 0 is $(1 - e^{-\frac{m}{2^r}})$

– The probability (p') that none of m distinct elements has tail length at least r is $(1 - 2^{-r})^m \simeq e^{-m2^{-r}}$

Therefore, we can observe that

@ if $2^r \gg m$, $p = \frac{m}{2^r} \to 0$ ; $p' = 1$

[1.5] In this case, the hashed result is never likely with R trailing 0s. so the R cannot be too large.

(b) if $2^r \ll m$, $p = 1 - e^{-m/2^r} \to 1$ ; $p' = 0$.

[1.5] In this case, every hashed result is likely with R trailing 0s. So the R cannot too small.

In general, R should be neither too large or too small. $2^R$ ~~should~~ should be around m.

4) Because the $k^{th}$ moment is very powerful and informative tool when we wish to measure some features about a stream.

→ $0^{th}$ moment : shows the number of distinct elements in the stream

→ $1^{th}$ moment : tells the length of the stream

→ $2^{th}$ moment : rep. the elements uneven of distribution in the S.

[1 POINT]