

Quiz7 Streaming

1)

Suppose we have t target and d dart ($t = n, d = km$)

Prob. of a target hit by a dart = $1/t$

Prob. of a target not hit by a dart = $1 - 1/t$

Prob. of a target not hit by all darts = $(1 - 1/t)^d$

(Since $(1 - 1/t)^t = 1/e$ for large t)

We have $(1 - 1/t)^{t \cdot d/t} = e^{-d/t}$

Prob. of a given target hit by at least one dart = $1 - e^{-d/t}$

Fraction of 1's = the probability that a bit in the array is set to 1 by at least one hashing

$$1 - e^{-km/n}$$

2) (refers to lecture "streaming" page 59-60)

Problems in combining estimates

- We can hash multiple times, take **avg.** of 2^R values
- Problem: ExpectedValue(2^R) $\rightarrow \infty$
 - When $2^R \geq m$, increase R by 1 \Rightarrow **probability halves, but value 2^R doubles**
 - $p = 1 - (1 - 2^{-R})^m = 1 - e^{-\frac{m}{2^R}} = \frac{m}{2^R}$, if $2^R \gg m$
 - Contribution from each large R to $E(2^R)$ grows, when R grows
 - Can have very large 2^R
- What about taking median instead?

• Avg only: what if **one very large value**?

• Median: all values are power of 2

– 1, 2, 4, 8, ..., 1024, 2048, ...

• Solution:

– Partition hash functions into **small groups**

– Take **average for each group**

– Take the **median of the averages**

3) (refers to lecture "streaming" page 68-70)

X1.element = c X1.value = 3

X2.element = d X2.value = 2

X3.element = a X3.value = 2

Estimate of 2^{nd} moment = **$n(2 \cdot X.\text{value} - 1)$**

Estimate using X_1 : $15(6-1) = 75$

Estimate using X_2 or X_3 : $15(4-1) = 45$

Avg. = $(75+45+45)/3 = 55$ (recall actual is 59)

4)

We want answer queries, how many times have we sold X in the last k sales (e.g., $k = 10, 20$, or 200 ; $N=100,000$)

How many 1s are in the last k bits? where $k \leq N$