Streaming Algorithms

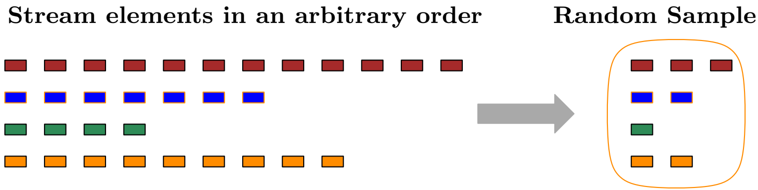
- Streaming Model of Computation
- Streaming Algorithms and DFA
- Stream: Motivation and Applications
- Synopsis: Sliding Window, Histogram, Wavelets
- Sampling from Stream: Reservoir Sampling
- Linear Sketch
- Count-Min Sketch
- AMS Sketch

Imdad ullah Khan

Stream Sampling

Synopsis: Random Sample

- Keep a “representative” subset of the stream
- Approximately compute query answer from sample (with appropriate scaling etc.)



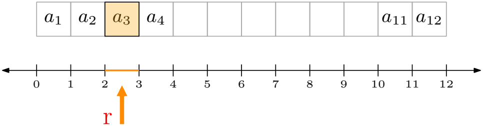
Stream elements in an arbitrary order

Random Sample

Sampling from an Array

Sample a random element from array A of length n ▷ $A[i]$ with prob $1/n$

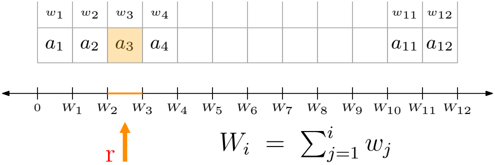
- Generate a random number $r \in [0, n]$ ▷ $r \leftarrow \text{RAND}() \times n$
- Return $A[\lceil r \rceil]$



Weighted Sampling from an Array

Sample random element (by weight) from array A $\triangleright$ $A[i]$ with prob. w_i/W

- Generate a random number $r \in [\,0, \sum_{j=1}^{n} w_i\,]$ $\triangleright$ $r \leftarrow \textsc{rand}() \times W_n$
- Return $A[\,i\,]$ if $W_{i-1} \leq r < W_i$



w_1 w_2 w_3 w_4 w_{11} w_{12}
a_1 a_2 a_3 a_4 a_{11} a_{12}
0 W_1 W_2 W_3 W_4 W_5 W_6 W_7 W_8 W_9 W_{10} W_{11} W_{12}
r

$$W_i = \sum_{j=1}^{i} w_j$$

Sampling an element from a Stream: Reservoir Sampling

Sample a random element from the stream S ▷ a_i with prob. $1/m$

- If m is known, use algorithm for sampling from array. For unknown m

```
Algorithm : Reservoir Sampling ($\mathcal{S}$)
    $R \leftarrow a_1$                                  ▷ $R$ (reservoir) maintains the sample
    for $i \geq 2$ do
        Pick $a_i$ with probability $1/i$
        Replace with current element in $R$
```

Prob. that a_i is in the sample R_m (m: stream length or query time)

$$= \underbrace{\text{Pr that } a_i \text{ was selected at time } i}_{\frac{1}{i}} \times \underbrace{\text{Pr that } a_i \text{ survived in } R \text{ until time } m}_{\prod_{j=i+1}^{m}\left(1-\frac{1}{j}\right)}$$

$$= \frac{1}{\cancel{i}} \times \frac{\cancel{i}}{\cancel{i+1}} \times \frac{\cancel{i+1}}{\cancel{i+2}} \times \frac{\cancel{i+2}}{\cancel{i+3}} \times \ldots \times \frac{\cancel{m-2}}{\cancel{m-1}} \times \frac{\cancel{m-1}}{m} = \frac{1}{m}$$

Sampling k elements from a Stream: Reservoir Sampling

Sample k random elements from the stream S ▷ a_i with prob. $^k/_m$

Algorithm : Reservoir Sampling $(\mathcal{S}, k)$

```
$R \leftarrow a_1, a_2, \ldots, a_k$                    ▷ $R$ (reservoir) maintains the sample
for  $i \geq k+1$ do
    Pick $a_i$ with probability $^k/_i$
    If $a_i$ is picked, replace with it a randomly chosen element in $R$
```

Prob. that a_i is in the sample R_m (m: stream length or query time)

$$= \underbrace{\text{Pr that } a_i \text{ was selected at time } i}_{\frac{k}{i}} \times \underbrace{\text{Pr that } a_i \text{ survived in } R \text{ untill time } m}_{\prod_{j=i+1}^{m}\left(1-\left(\frac{k}{j}\times\frac{1}{k}\right)\right)}$$

$$= \frac{k}{\cancel{i}} \times \frac{\cancel{i}}{\cancel{i+1}} \times \frac{\cancel{i+1}}{\cancel{i+2}} \times \frac{\cancel{i+2}}{\cancel{i+3}} \times \ \ldots \times \frac{\cancel{m-2}}{\cancel{m-1}} \times \frac{\cancel{m-1}}{m} = \frac{k}{m}$$