Describe one approximate algorithm for lossy counting &one for sketch

Lossy Counting and Sketch are both techniques used for approximate data stream processing, where the goal is to provide an approximate summary of the data in a memory-efficient manner.

Lossy Counting Algorithm:

Objective: Keep approximate counts of items in a data stream while using limited memory.

1. Initialization:

- Set parameters: `width` (bucket size) and `epsilon` (error threshold).

- Initialize a hash table with counters for each item.

2. Processing the Stream:

- Read each item from the stream.

- If the item is already in the hash table, increment its counter.

- If the item is not in the hash table, add it with a counter set to 1.

3. Periodic Deletion:

- After processing a certain number of items (`width`), scan the hash table.

- For each item, if its count is less than a threshold (`epsilon` times the number of processed items), remove it.

4. Result:

- The remaining items in the hash table with their counts are approximate frequent items in the stream.

Note: Lossy Counting sacrifices accuracy for memory efficiency, and it may not capture infrequent items well.

Sketch Algorithm (Count-Min Sketch):

Objective: Estimate frequencies of items in a data stream using limited memory.

1. Initialization:

- Set parameters: `d` (number of hash functions) and `w` (width of each hash table).

- Initialize a 2D array (counters) with dimensions `d x w` filled with zeros.

2. Processing the Stream:

- For each item in the stream, hash it using each of the `d` hash functions to determine `d` positions in the corresponding hash tables.

- Increment the counters at these positions.

3. Querying Frequency:

- To estimate the frequency of an item:

  - Hash the item using the same hash functions.

  - Retrieve the minimum value among the counters at the corresponding positions.

4. Result:

- The counters provide an approximate frequency for each item in the stream.

Note: Count-Min Sketch has the property of always providing a frequency estimate equal to or greater than the true frequency. It introduces a controlled level of error for memory efficiency. Adjusting

parameters `d` and `w` allows trade-offs between accuracy and memory usage.