PAREENITA A.SHIRSATH B.E.A.I.&.D.S. ROLL.NO : 57

BDA EXPERIMENT NO : 08

Flajolet-Martin (FM) Algorithm

```
import random
import math

def trailing_zeros(x):
    """ Counting number of trailing zeros
    in the binary representation of x."""
    if x == 0:
        return 1
    count = 0
    while x & 1 == 0:
        count += 1
        x >>= 1
    return count

def flajolet_martin(dataset, k):
    """Number of distinct elements using
    the Flajolet-Martin Algorithm."""
    max_zeros= 0
    for i in range(k):
        hash_vals = [trailing_zeros(random.choice(dataset))
                     for _ in range(len(dataset))]
        max_zeros = max(max_zeros, max(hash_vals))

    return 2 ** max_zeros

# Example
dataset = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
dist_num = flajolet_martin(dataset, 10)
print("Estimated number of distinct elements:", dist_num)
```

```
Estimated number of distinct elements: 8
```

```
import random
import hashlib

def hash_function(x, a, b, c):
    """Simple custom hash function."""
    hash_val = (a * hash(x) + b) % c
    return hash_val

def trailing_zeros(n):
    """Count number of trailing zeros in binary representation."""
    if n == 0:
        return 32  # large number if all zeros
    count = 0
    while (n & 1) == 0:
        n >>= 1
        count += 1
    return count

def flajolet_martin(stream, num_hash):
    print(f"Enter the space-separated values of a, b and c for {num_hash} hash functions:")
    hash_params = []
    for i in range(num_hash):
        a, b, c = map(int, input(f"Hash {i+1}: ").split())
        hash_params.append((a, b, c))

    max_trailing = [0] * num_hash

    for val in stream:
        for i, (a, b, c) in enumerate(hash_params):
            hv = hash_function(val, a, b, c)
            zeros = trailing_zeros(hv)
            max_trailing[i] = max(max_trailing[i], zeros)

    estimates = [2 ** r for r in max_trailing]
    avg_estimate = sum(estimates) / len(estimates)

    print(f"\nCounts for each hash: {estimates}")
    print(f"Average (Estimated distinct elements): {avg_estimate:.2f}")


# ----------- MAIN PROGRAM -----------
if __name__ == "__main__":
```

```
    n = int(input("Enter number of elements in the stream: "))
    print("Enter the elements (space-separated):")
    data_stream = input().split()

    num_hash = int(input("Enter the number of hash functions: "))
    flajolet_martin(data_stream, num_hash)
```

```
Enter number of elements in the stream: 10
Enter the elements (space-separated):
a b c a d  e f g h i
Enter the number of hash functions: 5
Enter the space-separated values of a, b and c for 5 hash functions:
Hash 1: 1 4 3
Hash 2: 2 6 7
Hash 3: 1 8 2
Hash 4: 3 4 7
Hash 5: 6 5 1

Counts for each hash: [4294967296, 4294967296, 4294967296, 4, 4294967296]
Average (Estimated distinct elements): 3435973837.60
```