Com S 435 PA4 Report

Jialin Xing

1. Count Min Sketch:

I imported an integer list with . There are should be 500,000 numbers with value 2;40 numbers with value 4; 1 number with value 10. I ran the program with epsilon 0.1, delta 0.5, q 0.2, r 0.1.

F(2) = 500002 with error: 0.000004

F(4) = 41 with error: 0.025

F(10) = 2 with error: 0.5

Therefore, high frequencies are well approximated.

In the same integer list, the HH returned the 2, 0 , 1, 3(which is the HH) and 9 other unique number which is not. HH does not performed good enough when q is valued just right q = r + epsilon.

1. Count Sketch:

Use same list:

F(2) = 500000

F(4) = 41 with error: 0.025

F(10) = 2 with error: 0.5

In conclusion, Count Sketch is more accurate than Count Min Sketch.