

General Instruction: After the fifth lecture of big data course, you should be able to answer the following questions.

1. Why does the Count-sketch generate the second hash function g_i to produce the random +1 and -1 values?
2. Consider two vectors (0,1,1) and (1,2,0). Please estimate the dot product of these two vectors using the Count-sketch algorithm with a 2×2 matrix. The hash functions g and h can be found as follows.

g hash functions:

	0	1	2
1	+1	-1	+1
2	-1	+1	-1

h functions

$$h1(j) = j \bmod 2$$

$$h2(j) = (j+1) \bmod 2$$

3. Consider a data stream with five items: 3, 1, 2, 1, 3. Please use FM sketch to estimate the number of distinct elements in this stream. Assume that the length of the bit array is 4 and the hash function $h(x) = (x * 11) \bmod 15$.