Quiz 3

Name:

(Estimating frequency moments of streams)

Problem: Consider a stream S = <a1, … , am> with elements from a domain D = {v1, v2, …, vn}. Let fi denote the frequency (or count) of a value vi in D.

The 2nd moment (F2) of the streams is

$$F_2 = \sum_{i=1}^{n} f_i^2$$

Estimate the F2 of the stream shown below using sketching algorithm.

Data stream is

6 6 10 9 4 5 8 7 5 6 10 4 9 6 7

Table Z:

	1	2	3	4	5	6	7	8	9	10
z1	1	1	-1	1	-1	1	1	-1	1	-1
z2	-1	1	1	1	1	1	1	-1	-1	1
z3	1	1	1	-1	-1	-1	-1	-1	1	1
z4	1	1	1	-1	1	-1	-1	-1	1	-1
z5	-1	-1	1	-1	-1	-1	-1	1	-1	-1

input	6	6	10	9	4	5	8	7	5	6	10	4	9	6	7
s1	1	2	1	2	3	2	1	2	1	1	0	1	2	3	4
s2	1	2	3	2	3	4	3	4	5	6	7	8	7	8	9
s3	-1	-2	-1	0	-1	-2	-3	-4	-5	-6	-5	-6	-5	-6	-7
s4	-1	-2	-3	-2	-3	-2	-3	-4	-3	-4	-5	-6	-5	-6	-7
s5	-1	-2	-3	-4	-5	-6	-5	-6	-7	-8	-9	-10	-11	-12	-13

➔ $(4^2 + 9^2 + (-7)^2 + (-7)^2 + (-13)^2) / 5 = 72.8$

Actual F_2:

	1	2	3	4	5	6	7	8	9	10	F_2
freq	0	0	0	2	2	4	2	1	2	2	37

In the slide,

$$\mathbb{E}\left[s^2\right] = \mathbb{E}\left[\sum_{i,j\in[n]} z_i z_j f_i f_j\right] = \sum_{i,j\in[n]} f_i f_j \mathbb{E}\left[z_i z_j\right] = \sum_{i\in[n]} f_i^2$$

Why the first equation

$$\mathbb{E}\left[s^2\right] = \mathbb{E}\left[\sum_{i,j\in[n]} z_i z_j f_i f_j\right]$$

Hold?

➔ $s = \sum_i z_i f_i$

Thus, $s^2 = (\sum_i z_i f_i)(\sum_j z_j f_j) = \sum_i \sum_j z_i z_j f_i f_j$