Assignment 4

Shown below is a data stream with N = 22 and the current bucket configuration. New elements enter the window at the right. Thus, the oldest bit of the window is the leftmost bit shown.

[1 0 1 1 0 0 0 1] 0 [1 1 1 0 1] [1 0 0 1] 0 [1] [1] 0

1. What is the largest possible bucket size for N = 22?

Answer: largest possible bucket size is $2^{\log N} = 16$

2. What is the estimate of the number of 1's in the latest k = 15 bits of this window?

Answer: Estimated number of 1's in latest k = 15 bits = sum of all the non-overlapping buckets before k + ½*total number of 1 in overlapping bucket = 1+1+2+4+(1/2)*4 = 10

3. The following bits enter the window, one at a time: 1 0 1 1 1 0 0 1. What is the bucket configuration in the window after this sequence of bits has been processed by DGIM?

Answer: Bucket composition after new bits enter the stream:

[1 0 1 1 0 0 0 1 0 1 1 1 0 1] [1 0 0 1 0 1 1] 0 [1 0 1] [1 1] 0 0 [1]

4. After having processed the bits from (3), what is now the estimate of the number of 1's in the latest k = 15 bits of the window?

Answer: Estimated number of 1's in latest k = 15 bits = sum of all the non-overlapping buckets before k + ½*total number of 1 in overlapping bucket = 1+2+2+(1/2)*4 = 7

5. In the file extension_DGIM.pdf you find 2 slides that explain how to generalize the DGIM algorithm from a bit stream to positive integers. Analogously to the slide example, work out the bit streams for the following stream of 8 numbers (oldest first): (125, 2, 77, 5, 13, 9, 99, 56). Compute the result for k = 3.

Answer: Each column represents the number in 8 bits:

	125	2	77	5	13	9	99	56
C_6	[1	0	1]	0	0	0	[1]	0
C_5	[1	0	0	0	0	0	1]	[1]
C_4	[1]	0	0	0	0	0	0	[1]
C_3	[1	0	1]	0	[1	1]	0	[1]
C_2	[1	0	1]	[1]	[1]	0	0	0
C_1	0	[1]	0	0	0	0	[1]	0
C_0	[1	0	1]	[1	1]	[1]	[1]	0

We estimate for 7 stream each representing a bit of the number, Ci represents number of 1's in row C_i for latest k bits

$C_0 = 1+1 = 2$, $C_1 = 1$, $C_2 = 0$, $C_3 = 1+ (1/2)*2 = 2$, $C_4 = 1$, $C_5 = 1+ (1/2)*2 = 2$, $C_6 = 1$

Total sum = $C_6*2^6+ C_5*2^5+ C_4*2^4+ C_3*2^3+ C_2*2^2+ C_1*2^1+ C_0*2^0 = 164$ = actual value (9+99+56)