Exercise 8.4

zaterdag 7 oktober 2023 16:37

The claim is false. To see why, consider the following stream for $\epsilon = \frac{1}{3}$: $\sigma = \langle 1, 1, 2, 2, 3, 3, 3, 3 \rangle$. Clearly, the only element which is ϵ-frequent in this case is 3. Furthermore, we have that $\frac{1}{\epsilon} = 3$; thus, we have that $|I| \geq \frac{1}{\epsilon}$ (and hence, that counters are decremented) when a third element is added to I. As a result, the algorithm will proceed as follows:

Stream element being processed	State of I after processing element, where (j, c) denotes an element and its counter value
start	$\emptyset$
1	$\{(1,1)\}$
1	$\{(1,2)\}$
2	$\{(1,2), (2,1)\}$
2	$\{(1,2), (2,2)\}$
3	$\{(1,2), (2,2)\}$
3	$\{(1,2), (2,2)\}$
3	$\{(1,2), (2,2)\}$
3	$\{(1,2), (2,2)\}$

The reason why the algorithm does not add the element 3 to its elements is that, whenever it is being added, it is the only element which has counter value 1; this element will then be the only element to be (immediately) removed from the set I, and it will never stay in the set. Thus, we see that decrementing the counter only when it has value 1 will lead to incorrect solutions, and hence, the claim must be false.

(More generally: as soon as all elements in the set I have a counter value of at least two, no element in the set can be removed from it *(apart from a newly added element being removed immediately after being added)*; if an ϵ-frequent item appears for the first time after this happens, then it will never stay in the set I and will hence not be reported in the end.)