

## Special case of most frequent item in a stream

Suppose to have a stream of  $n$  items, so that one of them occurs  $> n/2$  times in the stream. Also, the main memory is limited to keeping just two items and their counters, plus the knowledge of the value of  $n$  beforehand. Show how to find deterministically the most frequent item in this scenario. [Hint: since the problem cannot be solved deterministically if the most frequent item occurs  $\leq n/2$  times, the fact that the frequency is  $> n/2$  should be exploited.]

### SOLUTION

This solution works with just one counter. We initialize the counter with zero, then we do the following:

- add the first element of the stream to the counter with value 1
- if the next element is the same of the one stored by the counter, then we increment the counter by 1.
- if the element is different, we decrease the counter by 1, but if the counter become zero then we substitute the counter item with the new element, and we set the counter to zero.

Now, we give a little proof of correctness. Suppose we  $x_1, \dots, x_n$  element in the stream, with their relative frequency  $f_1, \dots, f_n$ , and suppose the element element with maximum frequency is  $x_{i^*}$ . Therefore by hypothesis we have:

$$\sum_{\forall i, i \neq i^*}^n f_i < \frac{n}{2} < f_{i^*}$$

Then by the following mathematical steps we have:

$$\begin{aligned} \sum_{\forall i, i \neq i^*}^n f_i &< f_{i^*} \\ \sum_{\forall i, i \neq i^*}^n f_i - f_{i^*} &< 0 \\ f_{i^*} - \sum_{\forall i, i \neq i^*}^n f_i &> 0 \end{aligned}$$

Notice that this inequality is true by hypothesis. Thus the element remaining in the counter must be the one with higher frequency.