Exercise 10.5

Saturday, 14 October 2023 12:18

Consider the following stream consisting of m items, where the last $\left\lceil \frac{m}{2} \right\rceil$ items are all zeroes:
$\left\langle 0,1,2,3, \dots, \left\lfloor \frac{m}{2} \right\rfloor, 0,0,0, \dots, 0 \right\rangle$

This stream is such that the number of zeroes is given by $1 + m - \left\lfloor \frac{m}{2} \right\rfloor = 1 + \frac{m}{2} + \left(\frac{m}{2} \bmod 1 \right) \geq 1 + \frac{m}{2}$. Given that we have $k = 10$, we have 10 groups. Furthermore, given that we have $t = 1$, we compute the minimum over only a single 'parallel iteration' of the algorithm; this means that we can ignore the effect of the min-trick. The item 0, which has frequency at least $\frac{m}{2} + 1$, will be assigned to one of the groups. By construction, we have that the group which contains this element will have that all other elements in this group will have their frequency overestimated by at least $\frac{m}{2} + 1$. Thus, if this group contains at least one other element, then we have an example in which the probability of overestimating the frequency of at least one item is at least 0.99.

It remains to show how we can achieve this probability. We note that the number of distinct elements which are unequal to 0 in the stream is $\left\lfloor \frac{m}{2} \right\rfloor - 1$. The only way for there to be no item among those $\left\lfloor \frac{m}{2} \right\rfloor - 1$ which does not map to the same group as 0, is for all of them to map to a different group. Since the hash function is uniform random, we have the probability for an item to map to a different group than the item 0 is $\frac{9}{10}$. Then, the probability for all $\left\lfloor \frac{m}{2} \right\rfloor - 1$ items to map to a different group is given by $\left(\frac{9}{10}\right)^{\left\lfloor \frac{m}{2} \right\rfloor - 1}$. This gives that the probability for there to be at least one item which maps to the same group as the item 0 is $1 - \left(\frac{9}{10}\right)^{\left\lfloor \frac{m}{2} \right\rfloor - 1}$, which we want to be at least 0.99. Hence, we obtain

$$1 - \left(\frac{9}{10}\right)^{\left\lfloor \frac{m}{2} \right\rfloor - 1} \geq 0.99$$

$$0.01 \geq \left(\frac{9}{10}\right)^{\left\lfloor \frac{m}{2} \right\rfloor - 1}$$

$$0.01 \cdot \left(\frac{9}{10}\right) \geq \left(\frac{9}{10}\right)^{\left\lfloor \frac{m}{2} \right\rfloor}$$

$$\log_{\frac{9}{10}} \left(0.01 \cdot \frac{9}{10}\right) \geq \log_{\frac{9}{10}} \left(\frac{9}{10}\right)^{\left\lfloor \frac{m}{2} \right\rfloor}$$

$$\left(\log_{\frac{9}{10}} 0.01\right) + 1 \geq \left\lfloor \frac{m}{2} \right\rfloor$$

Since we have that the logarithm is approximately 44, we have that the inequality must be satisfied for (among other values) $m \geq 100$. Hence, by picking a sufficiently large m, it is indeed possible to obtain a stream where the overestimation of at least one item is at least $\frac{m}{2}$ with a probability of at least 0.99.