Solutions for exercise sheet 2

Solution 1.
See code linked from the unit web page.

Solution 2.
See code linked from the unit web page.

Solution 3.
Recall that items are deleted in groups of k in the conceptual version of the algorithm described in the lecture notes. This tells us the estimate for an item can be at most m/k smaller than the true count. To improve this analysis observe that those counts currently in the data structure correspond to items in the stream whose counts have not been yet been deleted. So there are at most $m - \hat{m}$ items that could have been involved in deletions so far. Therefore an item can be at most $(m - \hat{m})/k$ smaller than the true count.

Solution 4.
Choose $k = 2/\epsilon$, and output all items satisfying $\hat{f}_j \geq \frac{\epsilon}{2}m$.

The final estimates satisfy

$$f_j - \frac{\epsilon}{2}m \leq \hat{f}_j \leq f_j.$$

If $j \in \mathrm{HH}_\epsilon(\sigma)$ then $\hat{f}_j \geq \epsilon m - \frac{\epsilon}{2}m = \frac{\epsilon}{2}m$, and so we output j. Conversely, if we output j then $f_j \geq \hat{f}_j \geq \frac{\epsilon}{2}m$, and so $j \in \mathrm{HH}_{\epsilon/2}(\sigma)$.

Solution 5.
The number of counters remaining is at most k because there are at most k counters with a larger count than the $(k+1)$th taken from largest to smallest. To compute the bounds we observe that counts can be lost in three places. First in the running of Misra-Gries on the first stream, second in the running of Misra-Gries on the second stream and third in the final pruning step. We know that an individual counter has been reduced by at most $\frac{m_1 - m_1'}{k}$ in the first stream and $\frac{m_2 - m_2'}{k}$ in the second. For the third step we need to show that the additional error C_{k+1} is such that

$$C_{k+1} \leq (m_1' + m_2' - m_{12}')/k$$

where m_{12}' is the sum of the counters after the final reduce step. If there were no more than k counters after the combine step then $C_{k+1} = 0$. Otherwise the prune step reduces the sum of the counters by at least

$(k+1)C_{k+1}$, So we have $m'_{12} \leq m'_1 + m'_2 - (k+1)C_{k+1}$ and the inequality follows.

Solution 6.
See code linked from the unit web page.

Solution 7.
No solution given.