1. Suppose we have run the (one-pass) Misra-Gries algorithm on two streams σ_1 and σ_2, thereby obtaining a summary for each stream consisting of k counters. Consider the following algorithm for merging these two summaries to produce a single k-counter summary.

   - Combine the two sets of counters, adding up counts for any common items.
   - If more than k counters remain:
     - $c \Leftarrow$ value of $(k+1)$th counter, based on decreasing order of value.
     - Reduce each counter by c and delete all keys with non-positive counters.

   Prove that the resulting summary is good for the combined stream $\sigma_1 o \sigma_2$ (here o denotes concatenation of streams) in the sense that frequency estimates obtained from it satisfy the bounds given below:

$$f_j - m/k \leqslant \hat{f}_j \leqslant f_j$$

2. Let $\sigma = \langle a_1, \ldots, a_m \rangle$ be a stream of m distinct items in the stream model. We wish to compute an element of rank $m/4$ in σ. Since this is hard to do exactly, we are satisfied with an item a_i such that $m/8 \leqslant \mathit{rank}(a_i) \leqslant 3m/8$. Present a stream algorithm that, for a given value $\delta > 0$, returns an item whose rank lies in the correct range with probability at least $1 - \delta$, and analyze the storage requirements of your algorithm. Note that $\mathit{rank}(a_i)$ is 1 plus the number of items in σ smaller than a_i.

3. Suppose we have a randomized streaming algorithm Alg whose goal is to estimate some function $\Phi(\sigma)$ of an input stream σ, where $\Phi(\sigma) > 3$. Let $B(n, m)$ be the number of bits of storage used by Alg, where m is the length and n is the size of the underlying universe. Suppose furthermore that we know that Alg outputs a value $\hat{\Phi}(\sigma)$ such that $\mathrm{E}(\hat{\Phi}(\sigma)) = \Phi(\sigma)$ and $\mathrm{Var}(\hat{\Phi}(\sigma)) = (1/3) \cdot \Phi(\sigma)$.

   - Show that there is a constant c with $0 < c < 1$ such that

$$\Pr(|\hat{\Phi}(\sigma) - \Phi(\sigma)| \geqslant c \cdot \Phi(\sigma)) \leqslant 1/6$$

   - Describe an randomized streamng algorithm (using Alg as a subroutine) that, given a parameter $\delta > 0$, computes an estimate $\hat{\Phi}(\sigma)$ such that $\Pr(|\hat{\Phi}(\sigma) - \Phi(\sigma)| \leqslant c \cdot \Phi(\sigma))$ with probability at least $1-\delta$ where c is the constant you determined in the above question. Prove that $\hat{\Phi}(\sigma)$ indeed has the desired accuracy with probability at least $1 - \delta$. Also analyze the amount of storage used by your algorithm.

4. Consider the following problems in the streaming model. Either prove that any deterministic streaming algorithm that solves these problems exactly must use $\Omega(m)$ bits in the worst case, or give a deterministic streaming algorithm that solves these problems using a sub-linear number of bits. If you give an algorithm, you should also prove its correctness and analyze the number of bits of storage it uses.

   - Given a stream $\sigma = a_1, \cdots, a_m$ over the universe $[n]$, with $m \leqslant n$, decide if all items in σ are distinct.

- Given a stream $\sigma = a_1, \cdots, a_m$ over the universe $[n]$, with $m = n - 2$ in which all items in σ are diferent, compute the items $j_1, j_2 \in [n]$ that are missing from σ. Note that only streams of length $n - 2$ are considered and that all items in the stream are distinct, which implies there are exactly two missing items.