Question 1. We want to design an algorithm that, given a list of distinct integers \( y_1, y_2, \ldots, y_n \) distinguishes between Case 1: The list is sorted, i.e., \( y_1 < y_2 < \ldots < y_n \) and Case 2: The list is \( \epsilon \)-far from being sorted, i.e., greater than \( \epsilon n \) elements need to be changed to make the list sorted. If the list is sorted, the algorithm should accept the input with probability 1 and if the list is \( \epsilon \)-far from being sorted the algorithm should reject with probability at least \( 3 \epsilon \).

1. Consider the algorithm that randomly samples \( i \in [n-1] \) and checks whether \( y_i < y_{i+1} \). Construct an input that is \( 1/2 \)-far from being sorted for which this approach would require \( \Omega(n) \) queries to determine the list isn’t sorted.
2. Consider the algorithm that randomly samples \( i, j \in [n] \) and checks whether \( y_i < y_j \). Construct an input that is \( 3/4 \)-far from being sorted for which this approach would require \( \Omega(n) \) queries to determine the list isn’t sorted.
3. Consider the algorithm that random samples \( i \in [n] \); queries \( y_i \) and sets \( z \) to be this value; and then performs a binary search in the list for the value \( z \). If the values of the list queried in the search for \( z \) reveal that the list isn’t sorted, reject the output. Prove that repeating this process \( O(1/\epsilon) \) times and accepting the input if none of the binary searches reject the input yields an efficient tester. **Hint:** Call index \( i \) good if the search for \( y_i \) doesn’t lead to rejection and show that the values at good indices are sorted.

Question 2. TBA.

Question 3. Two rooted trees \( T_1 \) and \( T_2 \) are said to be isomorphic if there exists a bijection from the vertices of \( T_1 \) to the vertices of \( T_2 \) satisfying the following condition: for each internal vertex \( v \) of \( T_1 \), if \( v_1, v_2, \ldots, v_k \) are children of \( v \) in \( T_1 \) then \( f(v_1), f(v_2), \ldots, f(v_k) \) are the children of \( f(v) \) in \( T_2 \). Note that no ordering is assumed on the children of any internal vertex.

1. Given a rooted tree \( T \), associate a multivariate polynomial \( P_v(x_0, x_1, x_2, \ldots) \) with each vertex \( v \) in the tree as follows: If \( v \) is a leaf let \( P_v = x_0 \) and if \( v \) has height \( h \geq 0 \) (i.e., the maximum distance from \( v \) to a descendent leaf is \( h \)) then let \( P_v \)

\[
(x_h - P_{v_1})(x_h - P_{v_2}) \ldots (x_h - P_{v_k}) .
\]

Prove that for any \( v \in T_1, w \in T_2 \), the subtree of \( T_1 \) rooted at \( v \) is isomorphic to the subtree of \( T_2 \) rooted at \( w \) iff \( P_v = P_w \).

2. Design and analyze an efficient randomized algorithm for testing whether \( T_1 \) and \( T_2 \) are isomorphic. **Hint:** You may wish to appeal to the Schwartz-Zippel Lemma (proved in COMPSCI 611) that states that if \( P_1(x_0, x_1, \ldots) \) and \( P_2(x_0, x_1, \ldots) \) are different multivariate polynomial of total degree \( d \) and \( S \) is a fixed set of values then,

\[
\mathbb{P}[P_1(r_0, r_1, \ldots) = P_2(r_0, r_1, \ldots)] \leq d/|S|
\]
where \( r_0, r_1, \ldots \) are chosen independently and uniformly from the set \( S \).

**Question 4.** Let \( p \) be a prime number and let \( a, b \) be chosen uniformly at random from \( \{0, 1, \ldots, p-1\} \). Define \( p \) random variables \( Y_0, \ldots, Y_{p-1} \) where \( Y_i = a i + b \mod p \). Prove that these random variables are pairwise independent, i.e., for \( i \neq j \) and all \( \alpha, \beta \in \{0, 1, \ldots, p-1\} \):

\[
P[Y_i = \alpha, Y_j = \beta] = P[Y_i = \alpha] P[Y_j = \beta].
\]

Prove or disprove that the random variables 3-wise independent, i.e., for distinct \( i, j, k \) and all \( \alpha, \beta, \gamma \in \{0, 1, \ldots, p-1\} \), is it true that \( P[Y_i = \alpha, Y_j = \beta, Y_k = \gamma] = P[Y_i = \alpha] P[Y_j = \beta] P[Y_k = \gamma] \)?

Let \( f \) be some function such that for all \( i \), \( f(Y_i) = 1 \) with probability \( q \) and \( f(Y_i) = 0 \) with probability \( 1 - q \). Let \( X = \sum_i f(Y_i) \). Prove the best upper bound you can on \( P[X = 0] \). **Hint:** Think Chebyshev.

**Question 5.** Consider a 2-wise random hash function \( h : [n] \to [w] \), i.e., each \( h(i) \) is distributed uniformly at random from \( [w] \) and \( h(i) \) and \( h(j) \) are independent if \( i \neq j \). Let \( c \) be a 4-wise random function \( c : [n] \to \{-1, 1\} \). Let \( f = (f_1, f_2, \ldots, f_n) \in \mathbb{R}^n \) and \( X = \sum_{i \in \mathbb{Z^w}} (\sum_{j \in [n]: h(j) = i} f_j c(j))^2 \).

Prove that,

1. \( E[X] = \sum_{i \in [n]} f_i^2 =: F_2 \)
2. \( \forall [X] = O(F_2^2 / w) \)

Let \( \langle a_1, \ldots, a_m \rangle \) be a stream where each \( a_i \in [n] \) and define \( f_i = |\{j : a_j = i\}| \). Design a small space data stream algorithm that approximates \( F_2 \) such that the estimate \( \tilde{F}_2 \) satisfies

\[
P[|\tilde{F}_2 - F_2| \leq \epsilon F_2] \geq 1 - \delta.
\]