1. (2 marks) For a stream of inputs, recall the basic estimator to calculate the F_2 value of the stream:
   1. Let h : U → {−1, 1} be drawn uniformly at random from a strongly k-wise independent family of hash functions.
   2. Initialize Z = 0.
   3. For each stream element x_i ∈ U, Z ← Z + h(x_i)
   4. Return Z^2.

   In the above example, k is a placeholder for a number.

   To show that this basic estimator is unbiased, we need to show that E[Z^2] = F_2. What value of k is sufficient to show this?
   (a) 1
   (b) 2
   (c) 3
   (d) 4

   Solution: (b) 2

2. (2 marks) In question 1, we also need to show that Variance[Z^2] ≤ 2F_2^2. What value of k is sufficient to show this?
   (a) 1
   (b) 2
   (c) 3
   (d) 4

   Solution: (d) 4

3. (1 mark) In question 1, we can improve the accuracy of our estimator by using t hash functions h_1, h_2, ..., h_t instead of just one. If each estimator outputs Z^2_1, Z^2_2, ..., Z^2_t respectively, what is the relationship between the new estimator Y and Z^2_1, Z^2_2, ..., Z^2_t that gives us the correct output?
   (a) Y = t(Z^2_1 + Z^2_2 + ... + Z^2_t)
   (b) Y = Z^2_1 + Z^2_2 + ... + Z^2_t
   (c) Y = \frac{1}{t}(Z^2_1 + Z^2_2 + ... + Z^2_t)
   (d) Y = \frac{1}{t^2}(Z^2_1 + Z^2_2 + ... + Z^2_t)

   Solution: (c) Y = \frac{1}{t}(Z^2_1 + Z^2_2 + ... + Z^2_t)
4. (1 mark) In question 4, what is the relationship between \( \text{Variance}[Y] \) and the variance of our original estimator, \( \text{Variance}[Z] \)?

(a) \( \text{Variance}[Y] = t(\text{Variance}[Z]^2) \)

(b) \( \text{Variance}[Y] = \text{Variance}[Z^2] \)

(c) \( \text{Variance}[Y] = \frac{1}{t}(\text{Variance}[Z^2]) \)

(d) \( \text{Variance}[Y] = \frac{1}{t^2}(\text{Variance}[Z^2]) \)

Solution: (c) \( \text{Variance}[Y] = \frac{1}{t}(\text{Variance}[Z^2]) \)

5. (2 marks) In question 5, applying Chebyshev’s inequality, what is an upper bound on the value of \( \Pr(|Y - E[Y]| \geq \epsilon F_2) \)?

(a) \( \frac{1}{t\epsilon^2} \)

(b) \( \frac{2}{t\epsilon^2} \)

(c) \( \frac{1}{2t\epsilon^2} \)

(d) \( \frac{1}{t^2\epsilon^2} \)

Solution: (b) \( \frac{2}{t\epsilon^2} \)

6. (1 mark) Let us consider the property testing algorithm to test whether a graph \( G(V,E) \) is connected or \( \epsilon \)-far from being connected. \( |V| = n \) and \( |E| = m \). Let a connected component of \( G \) be a set of nodes in \( V \) connected to each other but to no other nodes. If \( G \) is connected, then there is only one connected component of \( G \).

![Connectivity Testing Algorithm](image)

When is \( G \) \( \epsilon \)-far from being connected?

(a) At least \( \epsilon |E| \) edges must be added to establish connectivity.

(b) Less than \( \epsilon |E| \) edges must be added to establish connectivity.

(c) At least \( \epsilon |V| \) edges must be added to establish connectivity.

(d) Less than \( \epsilon |V| \) edges must be added to establish connectivity.

Solution: (a) At least \( \epsilon |E| \) edges must be added to establish connectivity.

7. (1 mark) In question 6, is it possible for the given algorithm to reject \( G \) if it is connected?

(a) Yes

(b) No
8. (1 mark) In question 6, what is the running time of the algorithm?
   (a) $O(n^4 \epsilon^4 m^4)$
   (b) $O(n^3 \epsilon^3 m^3)$
   (c) $O(n^2 \epsilon^2 m^2)$
   (d) $O(n \epsilon m)$

   **Solution:** (b) $O(n^3 \epsilon^3 m^3)$

9. (1 mark) For question 6, consider the statement: “If $G$ is $\epsilon$-far from connected, then $G$ has more than $\epsilon m + 1$ connected components.” Is this statement true or false?
   (a) True
   (b) False

   **Solution:** (a) True

10. (1 mark) For question 6, consider the statement: “If $G$ is $\epsilon$-far from connected, then at most $\epsilon m^2$ connected components are small, i.e. have no more than $\frac{2m}{\epsilon}$ nodes.” Is this statement true or false?
    (a) True
    (b) False

   **Solution:** (b) False

11. (1 mark) Consider the enforce and test approach for testing whether a given graph $G(V,E)$ is a biclique. Remember that a biclique is a graph where there exists a partition of vertices into $V_1$ and $V_2$ such that the edge set is exactly $V_1 \times V_2$. The algorithm is as follows:
    1. Pick vertex $v_0$ arbitrarily.
    2. For $i = 1$ to $2/\epsilon$
    3. Pick a pair of vertices $u_i$ and $v_i$ uniformly at random
    4. If biclique property violated, then REJECT.
    5. End For
    6. ACCEPT

    Which part of the algorithm constitutes the “test” part of the approach?
    (a) Line 1
    (b) Line 4
    (c) Lines 1 to 6
    (d) Lines 2 to 6

   **Solution:** (d) Lines 2 to 6

12. (1 mark) In question 11, is it possible for the algorithm to reject $G$ if it is a biclique?
    (a) Yes
    (b) No

Name: ___________________________ Roll No.: ___________________________
13. (1 mark) In question [11], how can the biclique property be violated?
   (a) Some edges are missing.
   (b) Some edges are additional.
   (c) Some edges are missing and other edges are additional.

   **Solution:** (c) Some edges are missing and other edges are additional.

14. (2 marks) In question [11], if the graph $G$ is $\epsilon$-far from being a biclique, then what is the probability that a pair chosen in Line 3 violates the biclique property?
   (a) $\geq \frac{1}{\epsilon}$
   (b) $\geq \frac{\epsilon}{n^2}$
   (c) $\geq \epsilon$
   (d) $\geq \frac{\epsilon}{n}$

   **Solution:** (c) $\geq \epsilon$

15. (2 marks) In question [11], what is the probability that no violating pair is chosen over the course of the algorithm?
   (a) $\geq \frac{1}{\epsilon}$
   (b) $\leq \frac{\epsilon}{n^2}$
   (c) $\geq \frac{1}{n^2}$
   (d) $\leq \frac{1}{\epsilon^2}$

   **Solution:** (d) $\leq \frac{1}{\epsilon^2}$