Rules:

• Collaborations are allowed on the homework but answers must be written independently.
• Please write on your solution who you collaborated with.
• Solutions must be typeset in LaTeX.
• Email solutions to mcgregor@cs.umass.edu with subject “711 Homework 3 Solutions.”
• If a solution is late by \( h \geq 0 \) hours, marks will be scaled by a factor \( 0.75^{h/24} \).

Question 1 (10 marks). Prove or disprove the following statements:

1. The cover time of a random walk on an undirected connected graph is polynomial in the number of nodes in the graph.
2. The cover time of a random walk on a directed strongly-connected graph is polynomial in the number of nodes in the graph.
3. Adding an undirected edge to an undirected graph can increase the cover time of a random walk on the graph.

Question 2 (10 marks). Recall that a sequence of random variables \( Z_0, Z_1, \ldots \) is a martingale with respect to sequence \( X_0, X_1, \ldots \) if for all \( n \geq 0 \):

   1. \( Z_n \) is a function of \( X_0, X_1, \ldots, X_n \);
   2. \( \mathbb{E}[|Z_n|] < \infty \);
   3. \( \mathbb{E}[Z_{n+1}|X_0, \ldots, X_n] = Z_n \)

   and that \( Z_0, Z_1, \ldots \) is a martingale if it is a martingale with respect to itself. Prove or disprove the following:

   1. If \( Z_0, Z_1, \ldots \) is a martingale with respect to \( X_0, X_1, \ldots \), then it is also a martingale with respect to itself.
   2. If \( Z_0, Z_1, \ldots \) is a martingale with respect to itself and \( Z_n \) is a function of \( X_0, X_1, \ldots, X_n \) (for all \( n \geq 0 \)) then \( Z_0, Z_1, \ldots \), is martingale with respect to \( X_0, X_1, \ldots \).

Question 3 (10 marks). Consider the following strange algorithm for finding the median of a set \( X = \{x_1, x_2, \ldots, x_{2n-1}\} \) of distinct integers.

1. Let \( \pi \) be a permutation of \([2n-1]\) chosen uniformly at random and let \( y_i = x_{\pi(i)} \)
2. Let \( X = \{y_1, \ldots, y_s\} \) and set \( L = 0 \) and \( H = 0 \)
3. For \( i = s+1, \ldots, 2n-1 \):
   a. If \( y_i < \min X \): \( L \leftarrow L + 1 \)
   b. If \( y_i > \max X \): \( H \leftarrow H + 1 \)
   c. If \( \min X < y_i < \max X \): Add \( y_i \) to \( X \)
      i. If \( H < L \): Remove largest element from \( X \), and \( H \leftarrow H + 1 \)
      ii. Else: Remove smallest element from \( X \), and \( L \leftarrow L + 1 \)
   d. If \( 1 \leq n - L \leq s \) then return the \((n - L)\)-th smallest element in \( X \); Otherwise “Fail”

How large must \( s \) be such that the above algorithm correctly returns the median of \( X \) with probability at least \( 9/10 \)?

Question 4 (10 marks). Consider a 2-wise random hash function \( h : [n] \rightarrow [w] \), 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
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 $X = \sum_{i \in [n]} (\sum_{j \in [n]} h(j) = i) f_j c(j)^2$. Prove that,

1. $\mathbb{E}[X] = \sum_{i \in [n]} f_i^2 =: F^2$
2. $\mathbb{V}[X] = O(F^2 / w)$

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 space stream algorithm that approximates $F_2$ such that the estimate $\hat{F}_2$ satisfies

$$
\mathbb{P}\left[ \left| \hat{F}_2 - F_2 \right| \leq \epsilon F_2 \right] \leq 1 - \delta.
$$

**Question 5** (10 marks). Let $\langle a_1, \ldots, a_m \rangle$ be a stream where each $a_i \in [n]$. Define $f_i = \{ j : a_j = i \}$ and $p_i = f_i / m$. Suppose all $p_i \leq 1/10$. Based on the algorithm for estimating $F_k = \sum_{i=1}^n f_i^k$ ($k \geq 3$) by Alon, Matias, Szegedy (JCSS 1999), design a data stream algorithm using $\tilde{O}(\epsilon^{-2} \log \delta^{-1} \log m (\log m + \log n))$ space that estimates the entropy,

$$
H(p) = -\sum_{i=1}^n p_i \ln p_i
$$

such that the estimate $\hat{H}$ satisfies

$$
\mathbb{P}\left[ \left| \hat{H} - H(p) \right| \leq \epsilon H(p) \right] \leq 1 - \delta.
$$