There are 2 questions for a total of 50 points.

---

1. (15 points) In class we saw a construction of a 4-wise independent hash function family. A similar notion is that of a 4-wise independent subset $S \subseteq \{0,1\}^n$. A subset S of n-bit strings is called 4-wise independent if for any four distinct indices $1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq n$ and a uniformly chosen string $x \in S$, $(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4})$ is distributed uniformly over $\{0,1\}^4$.

   Give a construction for such a 4-wise independent set over $\{0,1\}^n$. Give reasons. What is the size of S in your construction? You may use assumptions and results done in class.

2. (35 points) We saw some space lower bound arguments for deterministic streaming algorithms in the class. Here, you will be asked to develop ideas for showing space lower bounds for randomized streaming algorithms using results from randomized *communication complexity*. Let me first discuss communication complexity.

   Communication complexity: In most general setting there are two parties Alice and Bob. Alice holds a string $x \in \{0,1\}^n$ and Bob holds a string $y \in \{0,1\}^n$. Alice wants to send a communication to Bob so that he is able to compute a function $f(x,y)$. The simplest way for Alice is to just send x and Bob computes $f(x,y)$. However, the *communication complexity* of this protocol is n bits. Alice wants to minimize the communication in terms of the number of bits. Alice and Bob are even allowed to execute a randomized communication protocol for this purpose. For an accuracy parameter δ. The randomized communication complexity for computing f, denoted by $\mathcal{R}^f_\delta$, is defined to be the minimum number of bits that need to be sent so that Bob is able to compute $f(x,y)$ with probability at least $(1-\delta)$.

   For certain interesting functions the randomized communication complexity is known. One such function is the disjoint function $DISJ_n(x,y)$ which is 1 if there does not exist an index i such that x_i and y_i are both 1 and 0 otherwise. The following communication complexity result is known for $DISJ_n$:

   Claim: $\mathcal{R}^{DISJ_n}_{1/3} = \Omega(n)$.

   Using the above communication complexity result you need to argue that any randomized streaming algorithm that outputs the frequency of the most frequent element of a stream $a_1, ..., a_n$ (with each element from the set $\{1, ..., m\}$) within a multiplicative accuracy of (1 ± 0.1) with probability at least $2/3$, uses at least $\Omega(\min(m,n))$ space.