1. Let \( x \in \mathbb{R}^n \) be the frequency vector of a stream of \( m \) items (insertions only).

Show how to use the CountMin+ sketch seen in class (for \( \ell_1 \) point queries) to estimate the median item in the stream in the following sense: assuming there is \( j^* \in [n] \) such that \( \sum_{i=1}^{j^*} x_i = \frac{1}{2} m \), report an index \( j \in [n] \) that with high probability satisfies \( \sum_{i=1}^{j} x_i \in (\frac{1}{2} \pm \epsilon) m \).

2. Give a complete analysis of algorithm CountMin++ seen in class, for \( \ell_1 \) point query of a general frequency vector \( x \in \mathbb{R}^n \) (i.e., allowing negative entries), as follows.

   (a) Show for CountMin (the basic algorithm) that for every \( i \in [n] \),
   \[
   \Pr[|\bar{x}_i - x_i| \geq \alpha||x||_1] \leq \frac{1}{4}.
   \]
   Explain whether it is okay to use a 2-universal or pairwise independent hash function.

   (b) Show for algorithm CountMin++ (which runs \( k = O(\log n) \) copies of CountMin and reports their median) that for every \( i \in [n] \),
   \[
   \Pr[|\bar{x}_i - x_i| \geq \alpha||x||_1] \leq \frac{1}{n^2}.
   \]
   Hint: Define an indicator \( Y_l \) for the event that copy \( l \in [k] \) succeeds, then use one of the concentration bounds.

   (c) Conclude by stating explicitly the storage required by this algorithm, including storage of hash functions.

3. Let \( A \) be a 0-1 matrix of size \((2^k - 1) \times k\) whose rows \( A_i \) are exactly all the nonzero vectors in \( \{0, 1\}^k \). For a random \( p \in \{0, 1\}^k \), define \( h_p : [2^k - 1] \to \{0, 1\} \) by \( h_p(i) := (Ap)_i = \langle A_i, p \rangle \), where all operations are performed modulo 2.

   Prove that the family \( H = \{h_p : p \in \{0, 1\}^k\} \) is pairwise independent.

   Conclude by stating explicitly the performance of this construction (number of bits needed to store \( n = 2^k - 1 \) pairwise independent random bits \( h(1), \ldots, h(n) \)).