Problem 1

Let us take $\gamma < 1$, and define a sequence of random variables by

$$X_{n+1} := \begin{cases} X_n + 1 & \text{w.p. } (1 + \gamma)^{-X_n} \\ X_n & \text{otherwise} \end{cases}$$

We will calculate expectation and variance of $T_n := (1 + \gamma)^{X_n}$. First, let us consider conditional expectation

$$E[T_{n+1}|X_n] = E[(1 + \gamma)^{X_{n+1}}|X_n]$$

$$= (1 + \gamma)^{-X_n} (1 + \gamma)^{X_n+1} + (1 - (1 + \gamma)^{-X_n}) (1 + \gamma)^{X_n}$$

$$= \gamma + (1 + \gamma)^{X_n}$$

and hence

$$ET_n = \gamma + ET_{n-1} = \cdots = \gamma n + ET_0 = \gamma n + 1 \quad (1)$$

Calculation of a second moment is similar

$$E[T_{n+1}^2|X_n] = E[(1 + \gamma)^{2X_{n+1}}|X_n]$$

$$= (1 + \gamma)^{-X_n} (1 + \gamma)^{2X_n+2} + (1 - (1 + \gamma)^{-X_n}) (1 + \gamma)^{2X_n}$$

$$= (1 + \gamma)^{X_n+2} - (1 + \gamma)^{X_n} + (1 + \gamma)^{2X_n}$$

$$= (\gamma^2 + 2\gamma)(1 + \gamma)^{X_n} + (1 + \gamma)^{2X_n}$$

hence

$$ET_{n+1}^2 = (\gamma^2 + 2\gamma)E[(1 + \gamma)^{X_n}] + ET_n^2$$

$$= (\gamma^2 + 2\gamma)(1 + n\gamma) + ET_n^2$$

$$= \cdots$$

$$= (\gamma^2 + 2\gamma) \left( 1 + n + \frac{\gamma n(n+1)}{2} \right) + 1.$$

We observe now that in variance calculation (i.e. $ET_n^2 - (ET_n)^2$) the $n^2\gamma^2$ term cancels out, and we have

$$Var(T_n) \leq C(n\gamma^2 + n^2\gamma^3) \leq C\gamma ET_n^2$$

for some universal constant $C$. If we take $\gamma < \frac{\varepsilon^2}{24C}$, we have $Var(T_n) < \frac{\varepsilon^2}{12} ET_n^2$, and by Chebyshev inequality

$$P(|T_n - ET_n| > \frac{\varepsilon}{2} ET_n) < \frac{1}{3}$$
Now, let us take \( W_n := X_{n-1} \), so we have \( \mathbb{E} W_n = n \). For \( n = \Omega \left( \frac{1}{\varepsilon^2} \right) \), we know that \( |T_n - \mathbb{E} T_n| < \varepsilon \mathbb{E} T_n \) implies \( W_n = (1 \pm \varepsilon)n \). We can use \( W_n \) as an estimator for \( n \). If \( n < \frac{1}{\varepsilon^2} \), we can store it directly.

We will focus now on space complexity of the algorithm. For any \( \eta \), once \( X_n \geq \log_{1+\gamma}(\eta^{-1}n) = \frac{\log \eta^{-1}n}{\log(1+\gamma)} \), with probability at least \( 1 - \eta \) it will not increase ever again. To store such an \( X \), we need

\[
\log \frac{\log \eta^{-1}n}{\log(1+\gamma)} \leq \log \frac{1}{\varepsilon} + \log \log \eta^{-1}n
\]

bits of memory.

**Problem 2**

Take some constant \( D > 1 \). Consider a tree with degree \( \lceil n^{\frac{1}{D}} \rceil \), of depth \( D \) (so that the number of leaves is about \( n \)). On every level of the tree, we can use a CountMin sketch with approximation parameter \( \frac{\varepsilon}{4} \) and failure probability \( \frac{\delta}{D} \). We have \( D \) levels, so by union bound probability that any of those sketches fails is at most \( \delta \). If none of the sketches failed, we can recover \( \ell_1 \) heavy hitters in time \( O(Dn^{\frac{1}{D}} \varepsilon^{-1} \log \frac{Du}{\delta}) \) — in each level \( L \) we have at most \( \frac{\varepsilon}{4} \) heavy hitters, and to find them, it is enough to do PointQuery on all the children of heavy hitters from the previous layer — number of such children is \( \varepsilon^{-1}n^{\frac{1}{D}} \).

The update operation takes time \( O(D \log(nD)) = O(\log(nD)) \) — for each of \( D \) levels, we have to use Update operation of corresponding CountMin structure.

**Problem 3**

We consider matrix \( \Pi \in \mathbb{R}^{m \times n} \), with entries given by

\[
\Pi_{ij} = \sigma_{ij} \delta_{ij}
\]

where all \( \sigma_{ij} \) are i.i.d. random signs, and \( \delta_{ij} \in \{0, 1\} \), are such that for each column \( j \), there is exactly one row \( i \) (chosen uniformly at random) with \( \delta_{ij} = 1 \).
We wish to calculate $E\|\Pi x\|_2^2$ and $Var(\|\Pi x\|_2^2)$.

\[
E\|\Pi x\|_2^2 = E \sum_{r \leq m} \left( \sum_i \sigma_{ri} \delta_{ri} x_i \right)^2
\]
\[
= E \sum_{r \leq m} \sum_{i,j} \sigma_{ri} \sigma_{rj} \delta_{ri} \delta_{rj} x_i x_j
\]
\[
= \sum_{r \leq m} \left( \sum_i E \sigma_{ri}^2 \delta_{ri}^2 x_i^2 + \sum_{i \neq j} (E \sigma_{ri} \sigma_{rj} \delta_{ri} \delta_{rj}) x_i x_j \right)
\]
\[
= \sum_{r \leq m} \frac{1}{m} \left( \sum_i x_i^2 \right)
\]
\[
= \sum_{i} x_i^2
\]

For variance calculation, let us first expand $W := \|\Pi x\|_2^2 - E\|\Pi x\|_2^2$. We have

\[
\|\Pi x\|_2^2 - E\|\Pi x\|_2^2 = \sum_{r \leq m} \sum_{i,j} \delta_{ri} \delta_{rj} \sigma_{ri} \sigma_{rj} x_i x_j - \sum_i x_i^2
\]
\[
= \sum_{r \leq m} \sum_{i,j} \sigma_{ri} \sigma_{rj} \delta_{ri} \delta_{rj} x_i x_j - \sum i \left( \sum_{r \leq m} \delta_{ri}^2 \right) x_i^2
\]
\[
= \sum_{r \leq m} \sum_{i \neq j} \sigma_{ri} \sigma_{rj} \delta_{ri} \delta_{rj} x_i x_j
\]

Let us use $W_{r,i,j}$ to denote a single term in this sum, i.e. $W_{r,i,j} := \sigma_{ri} \sigma_{rj} \delta_{ri} \delta_{rj} x_i x_j$. Note that for $r_1 \neq r_2$ we have $EW_{r_1,i_1,j_1} W_{r_2,i_2,j_2} = 0$. Similarly, if $r_1 = r_2$ we need one of the following two conditions in order for $EW_{r_1,i_1,j_1} W_{r_2,i_2,j_2}$ to be nonzero: either $i_1 = i_2 \land j_1 = j_2$ or $i_1 = j_2 \land i_2 = j_1$. For such terms, we have

\[
EW_{r,i,j}^2 = E\sigma_{r,i}^2 \sigma_{r,j}^2 \delta_{ri}^2 \delta_{rj}^2 x_i^2 x_j^2 = \frac{1}{m^2} x_i^2 x_j^2
\]

and similarly $EW_{r,i,j} W_{r,j,i} = \frac{1}{m^2} x_i^2 x_j^2$. Finally, this yields

\[
E \left( \|\Pi x\|_2^2 - E\|\Pi x\|_2^2 \right)^2 = 2 \sum_{r \leq m} \sum_{i \neq j} \frac{2}{m^2} x_i^2 x_j^2 = \frac{2}{m} \sum_{i \neq j} x_i^2 x_j^2
\]
We have $Var \|\Pi x\|^2 \leq \frac{2}{m} E\|\Pi x\|^2$, hence, by Chebyshev inequality

$$P \left( \|\Pi x\|^2 - \|x\|^2 > \varepsilon \|x\|^2 \right) \leq \frac{2}{m \varepsilon^2}$$

If we take $m \geq \frac{6}{\varepsilon^2}$ this probability will be smaller than $\frac{1}{3}$, and we can amplify it to arbitrary small $\delta$ in the standard way, by taking median of $\log \frac{1}{\delta}$ independent such estimators.