Problem 1

Let us take $\gamma < 1$, and define a sequence of random variables by $X_0 := 0$,

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

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

$$\begin{aligned} \mathbb{E}[T_{n+1}|X_n] &= \mathbb{E}[(1+\gamma)^{X_{n+1}}|X_n] \\ &= (1+\gamma)^{-X_n}(1+\gamma)^{X_n+1} + \left(1-(1+\gamma)^{-X_n}\right)(1+\gamma)^{X_n} \\ &= \gamma + (1+\gamma)^{X_n} \end{aligned}$$

and hence

$$\mathbb{E}T_n = \gamma + \mathbb{E}T_{n-1} = \cdots = \gamma n + \mathbb{E}T_0 = \gamma n + 1 \tag{1}$$

Calculation of a second moment is similar

$$\begin{aligned} \mathbb{E}[T_{n+1}^2|X_n] &= \mathbb{E}[(1+\gamma)^{2X_{n+1}}|X_n] \\ &= (1+\gamma)^{-X_n}(1+\gamma)^{2X_n+2} + \left(1-(1+\gamma)^{-X_n}\right)(1+\gamma)^{2X_n} \\ &= (1+\gamma)^{X_n+2} - (1+\gamma)^{X_n} + (1+\gamma)^{2X_n} \\ &= (\gamma^2+2\gamma)(1+\gamma)^{X_n} + (1+\gamma)^{2X_n} \end{aligned}$$

hence

$$\begin{aligned} \mathbb{E}T_{n+1}^2 &= (\gamma^2+2\gamma)\mathbb{E}\left[(1+\gamma)^{X_n}\right] + \mathbb{E}T_n^2 \\ &= (\gamma^2+2\gamma)(1+n\gamma) + \mathbb{E}T_n^2 \\ &= \cdots \\ &= (\gamma^2+2\gamma)\left(1+n+\frac{\gamma n(n+1)}{2}\right) + 1. \end{aligned}$$

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

$$Var(T_n) \leq C(n\gamma^2 + n^2\gamma^3) \leq C\gamma\mathbb{E}T_n^2$$

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

$$\mathbb{P}(|T_n - \mathbb{E}T_n| > \frac{\varepsilon}{2}\mathbb{E}T_n) < \frac{1}{3}$$

Now, let us take $W_n := \frac{X_n - 1}{\gamma}$, so we have $\mathbb{E}W_n = n$. For $n = \Omega(\frac{1}{\varepsilon^2})$, 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 η, 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 δ. If none of the sketches failed, we can recover ℓ_1 heavy hitters in time $\mathcal{O}(Dn^{\frac{1}{D}}\varepsilon^{-1} \log \frac{Dn}{\delta})$ — in each level L we have at most $\frac{1}{\varepsilon}$ 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 $\mathcal{O}(D \log(\frac{nD}{\delta})) = \mathcal{O}(\log(\frac{n}{\delta}))$ – 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 σ_{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 $\mathbb{E}\|\Pi x\|_2^2$ and $Var(\|\Pi x\|_2^2)$.

$$\begin{aligned}
\mathbb{E}\|\Pi x\|_2^2 &= \mathbb{E}\sum_{r\le m}\left(\sum_i \sigma_{ri}\delta_{ri}x_i\right)^2 \\
&= \mathbb{E}\sum_{r\le m}\sum_{i,j}\sigma_{ri}\sigma_{rj}\delta_{ri}\delta_{rj}x_ix_j \\
&= \sum_{r\le m}\left(\sum_i \mathbb{E}\sigma_{ri}^2\delta_{ri}^2x_i^2 + \sum_{i\ne j}(\mathbb{E}\sigma_{ri}\sigma_{rj}\delta_{ri}\delta_{rj})x_ix_j\right) \\
&= \sum_{r\le m}\frac{1}{m}\left(\sum_i x_i^2\right) \\
&= \sum_i x_i^2
\end{aligned}$$

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

$$\begin{aligned}
\|\Pi x\|_2^2 - \mathbb{E}\|\Pi x\|_2^2 &= \sum_{r\le m}\sum_{i,j}\delta_{ri}\delta_{rj}\sigma_{ri}\sigma_{r,j}x_ix_j - \sum_i x_i^2 \\
&= \sum_{r\le m}\sum_{i,j}\sigma_{ri}\sigma_{rj}\delta_{ri}\delta_{rj}x_ix_j - \sum_i\left(\sum_{r\le m}\delta_{ri}^2\right)x_i^2 \\
&= \sum_{r\le m}\sum_{i\ne j}\sigma_{ri}\sigma_{rj}\delta_{ri}\delta_{rj}x_ix_j
\end{aligned}$$

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

$$\mathbb{E}W_{r,i,j}^2 = \mathbb{E}\sigma_{r,i}^2\sigma_{r,j}^2\delta_{ri}^2\delta_{rj}^2x_i^2x_j^2 = \frac{1}{m^2}x_i^2x_j^2$$

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

$$\mathbb{E}\left(\|\Pi x\|_2^2 - \mathbb{E}\|\Pi x\|_2^2\right)^2 = 2\sum_{r\le m}\sum_{i\ne j}\frac{2}{m^2}x_i^2x_j^2 = \frac{2}{m}\sum_{i\ne j}x_i^2x_j^2$$

We have $Var\,\|\Pi x\|_2^2 \leq \frac{2}{m}\mathbb{E}\|\Pi x\|_2^2$, hence, by Chebyshev inequality

$$\mathbb{P}\left(\left|\|\Pi x\|_2^2 - \|x\|_2^2\right| > \varepsilon\|x\|_2^2\right) \leq \frac{2}{m\varepsilon^2}$$

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 δ in the standard way, by taking median of $\log\frac{1}{\delta}$ independent such estimators.