Problem 1

Propose a modification of Morris' algorithm for approximate counting, which returns $(1+\varepsilon)$-approximation to a counter n, with probability $\frac{2}{3}$, while using optimal $\mathcal{O}(\log \frac{1}{\varepsilon} + \log\log n)$ bits of memory (as opposed to $\mathcal{O}(\frac{\log\log n}{\varepsilon})$).

Hint: Take $\gamma < 1$, and consider counter X — on update operation increment it by 1 with probability $(1+\gamma)^{-X}$ (as opposed to 2^{-X}).

Problem 2

CountMin sketch, as presented on the lecture, provides following guarantees: using $\varepsilon^{-1} \log \frac{n}{\delta}$ words of memory, it can supports sequence of updates of form $x_i \leftarrow x_i + \Delta$, such that at the end of the stream with probability $1 - \delta$, it holds that $\forall_i query(i) = x_i \pm \varepsilon \|x\|_1$. One can use it to solve the ℓ_1-heavy hitters problem using the same amount of memory, but with query time $\mathcal{O}(n \log \frac{n}{\delta})$. The ℓ_1-heavy hitters problem is to identify all indices i such that $x_i \geq \varepsilon \|x\|_1$, while not reporting any i with $|x_i| \leq \frac{\varepsilon}{2} \|x\|_1$

With *dyadic trick* on top of the CountMin sketches, in the regime of failure probability $\delta = \text{poly}(1/n)$, this query complexity can be reduced to $\mathcal{O}(\log^2 n)$, but with space complexity $\mathcal{O}(\varepsilon^{-1} \log^2 n)$ and update time $\mathcal{O}(\log^2 n)$.

Show that, by using trees of constant depth (and large degree) instead of binary trees, it is possible to solve ℓ_1 heavy hitters problem using $\mathcal{O}(\varepsilon^{-1} \log n)$ words of memory and $\mathcal{O}(\log n)$ update time, with answering queries in time $\mathcal{O}(\varepsilon^{-1} n^{0.01})$ — strictly better than CountMin sketch itself. **Hint:** You should first note that without loss of generality, ε can be assumed to be at least $1/n$.

Problem 3

Consider matrix $\Pi \in \mathbb{R}^{m \times n}$, with entries given by

$$\Pi_{ij} = \sigma_{ij} \delta_{ij}$$

where all σ_{ij} are 4-wise independent random signs, and $\delta_{ij} \in \{0, 1\}$, are such that for each column j, there is exactly one row i (uniform) with $\delta_{ij} = 1$.

Show that if $m = \Omega(\frac{1}{\varepsilon^2})$ then

$$\mathbb{P}\left(\left|\|\Pi x\|_2^2 - \|x\|_2^2\right| \geq \varepsilon \|x\|_2^2\right) < \frac{1}{3}$$

In particular, such a matrix yields a sketch for estimation of second moment of a frequency vector, with the same space complexity as AMS (namely $\mathcal{O}(\varepsilon^{-2}\log\frac{1}{\delta})$), but significantly faster update time – $\log\frac{1}{\delta}$ instead of $\varepsilon^{-2}\log\frac{1}{\delta}$.