Przemysław Uznański 04/11/2019

Algorithms for Big Data

Spring Semester 2022

Exercise Set 5

Exercise 1:
Show how to use $\mathsf{CountMin}$ to support *range queries*: $\mathsf{range}(a, b) = \sum_{i=a}^{b} x_i \pm \varepsilon|x|_1$. The space usage should be $\operatorname{poly}\log$ worse than in the vanilla $\mathsf{CountMin}$.
Hint: You can use a few $\mathsf{CountMin}$ structures to obtain data structure with error guarantee slightly worse than in $\mathsf{CountMin}$. You can later just take ϵ' to be a little smaller than ε to offset this.

Exercise 2: (Quantiles)
ϕ-quantile of a multiset of size n is $\phi \cdot n$-smallest element. ε-approximate ϕ-quantile is any element that is between $(\phi - \varepsilon)$-quantile and $(\phi + \varepsilon)$-quantile. Use previous exercise to build a sketch that allows a queries of form $\text{quantile}(\phi)$ for a fixed in advance ε.

Exercise 3: (Sketching for inner product, 2pts)
Let X be a $\mathsf{CountMin}$ sketch of vector x, and Y be a $\mathsf{CountMin}$ sketch of vector y, with x and y being non-negative. Both sketches are obtained used the same hashing. Our goal is to approximate $x \odot y = \sum_i x_i y_i$. Show that $\min_j \sum_k X[j][i] \cdot Y[j][i]$ estimates $x \odot y$ up to $\pm\varepsilon|x|_1|y|_1$ error.