# $22\mathbf{C}:231$ Design and Analysis of Algorithms
# Homework 5: Elaboration and Concepts

**Approximate Median:** Let $S = \{a_1, a_2, \ldots, a_m\}$ be a multiset of $m$ numbers. The median of $S$ is defined to be the $\lceil m/2 \rceil$'th smallest number in $S$. Since $S$ is a multiset, what we mean is that if we sort $S$ in non-decreasing order and write into an array $A[1..m]$, the median is the number stored in $A[\lceil m/2 \rceil]$. For $\varepsilon < 1/2$, an $\varepsilon$-approximate median is any element which is the $t$-th smallest number of $S$ for some

$$t \in [(\frac{1}{2} - \varepsilon)m, (\frac{1}{2} + \varepsilon)m].$$

For example, if $S = \{1, 2, \ldots, 100\}$, and $\varepsilon = 0.05$, then any number in the range $\{45, 46, \ldots, 55\}$ is an $\varepsilon$-approximate median.

Suppose we pick $K$ elements from $S$ uniformly at random from $S$, independently (with replacement). Let $y$ denote the median of the $K$ elements. The probability that $y$ is not an $\varepsilon$-approximate median is at most

$$\frac{2}{e^{2\varepsilon^2 K/3}}.$$

This is a consequence of the Chernoff bound, which we will discuss later in the course.

**Sen Slope:** In the homework, we are given a set

$$\{(i, f(i)) \mid 1 \leq i \leq N\}.$$

Let

$$S = \{\frac{f(j) - f(i)}{j - i} \mid 1 \leq i < j \leq N\}.$$

The Sen slope is defined to be the median of the set $S$. In the homework, it suffices to compute any $\varepsilon$-approximate median of $S$ using the above sampling scheme, for $\varepsilon = 0.05$. We want the failure probability to be at most $1/10^6$.