# Randomized Algorithms 2019A – Problem Set 3

Robert Krauthgamer and Moni Naor

Due: Jan. 17, 2019

1. Design a randomized algorithm that approximates the median of an input list $a_1, \ldots, a_n \in \mathbb{R}$ as follows. Given also $\varepsilon \in (0, \frac{1}{2})$, the algorithm should report a number $a_i$ whose *rank* (position in the sorted list of $a_i$'s) is in the range $(\frac{1}{2} \pm \varepsilon)n$. The running time should be $O(1/\varepsilon^2)$ for (constant) high probability of success.

   Hint: Sample elements at random and count how many samples have rank below $(\frac{1}{2} - \varepsilon)n$.

2. We saw in class a randomized algorithm (call it Algorithm $E+$) with success probability $3/4$ for counting DNF solutions. (Here, success means that the output is within factor $1 \pm \varepsilon$ of $|S|$, the number of satisfying assignments of the input DNF formula.)

   Consider now Algorithm $E{+}{+}$, which repeats that Algorithm $E+$ independently $t = O(\log \frac{1}{\delta})$ times and then reports the *median* of their outputs. Prove that Algorithm $E++$ has success probability $1 - \delta$ (again, success means that the output is within factor $1 \pm \varepsilon$ of $|S|$).

   Hint: Use Chernoff-Hoeffding bounds to count the number of "successes".

3. Suppose $X_1, \ldots, X_n$ are chosen independently and uniformly at random from $-1, 1$, and let $S_t = \sum_{i=1}^{t} X_i$ be its prefix sum for $t \in [n]$. Notice that $S_1, \ldots, S_n$ describes a random walk on $\mathbb{Z}$.

   (a) Show that for $M = 8 \log n$,

   $$\Pr\big[\max_{t\in[n]} S_t \geq M\sqrt{n}\big] \leq \tfrac{1}{4}. \tag{1}$$

   (b) Show that (1) above holds even for $M = O(1)$ by first proving that

   $$\forall b \in \mathbb{R}, \qquad \Pr\big[\max_{t\in[n]} S_t \geq b\big] \leq 2\Pr\big[S_n \geq b\big].$$

4. Recall the construction from class of a coreset $Y$ for 1-median of a set $X = \{X_1, \ldots, X_n\} \subset \mathbb{R}^d$ via Importance Sampling. Specifically, $Y$ is a multiset of $m \geq L\varepsilon^{-2} \log \frac{1}{\delta}$ points, each sampled iid from $X$ according to the distribution given by $q(x) = \frac{s(x)}{S(X)}$, and then reweighted to $w(x) = \frac{1}{m\ q(x)}$. Prove that

   $$\Pr[w(Y) \in (1 \pm \varepsilon)n] \geq 1 - \delta. \tag{2}$$

   Hint: Show that all $x \in X$ satisfy $\frac{1}{q(x)} \leq O(n)$, and then use a concentration bound.

**Extra credit:**

5. Prove the "last" case in the construction of a strong coreset $Y$ for the geometric median of $X$ via importance sampling (called Theorem 12 in class), as follows. Suppose $Y$ satisfies

$$w(Y) \in (1 \pm \varepsilon)n,$$

(which happens WHP by (2) above), and

$$\tfrac{1}{m} \sum_{y \in Y} \frac{\|y - c^*\|}{q(y)} \in (1 \pm \varepsilon) \sum_{x \in X} \|x - c^*\|,$$

(which happens WHP by Lemma 9 in class applied to the optimal median $c^*$). Show that

$$\forall c \notin B(c^*, \tfrac{1}{\varepsilon}\tfrac{\mathrm{OPT}}{n}), \qquad f(Y, c) \in (1 \pm O(\varepsilon))\ f(X, c),$$

where $f(Y, c) = \sum_{y \in Y} w(y)\|y - c\|$ (and similarly for $X$ but with unit weights).

Hint: Give upper and lower bounds that show that $f(Y, c) \approx n\|c - c^*\|$ and also $f(X, c) \approx n\|c - c^*\|$.