CSE 521: Design and Analysis of Algorithms **Fall 2020**

Problem Set 2

Deadline: Nov 1st (at 11:59 PM) in Canvas

Instructions

- You should think about each problem by yourself for at least an hour before choosing to collaborate with others.
- You are allowed to collaborate with fellow students taking the class in solving the problems (in groups of at most 2 people for each problem). But you **must** write your solution on your own.
- You are not allowed to search for answers or hints on the web. You are encouraged to contact the instructor or the TAs for a possible hint.
- You cannot collaborate on Extra credit problems
- Solutions typeset in LATEX are preferred.
- Feel free to use the Discussion Board or email the instructor or the TA if you have any questions or would like any clarifications about the problems.
- Please upload your solutions to Canvas. The solution to each problem **must** be uploaded separately.

In solving these assignments and any future assignment, feel free to use these approximations:

$$1 - x \approx e^{-x}, \qquad n! \approx (n/e)^n, \qquad \left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

1) Let U be a universe. A family of hash functions $\mathcal{H} := \{h : U \to \{-1, +1\}\}$ is a sketching hash family with error $\epsilon > 0$ if for any function $f : U \to \mathbb{R}$, which is not identically zero,

$$\mathbb{P}\left[\sum_{i \in U} f(i)h(i) = 0\right] \leq \epsilon.$$

The value $sk_h(f) := \sum_{i \in U} f(i)h(i)$ is called the *sketch* of f.

a) Prove that if $f, f' : U \to \mathbb{R}$ are different functions, then

$$\mathbb{P}\left[sk_h(f) = sk_h(f')\right] \leq \epsilon.$$

b) Suppose $\mathcal{H}$ is the family of *all* functions from U to $\{-1, +1\}$. Prove that it is a sketching family with error $\epsilon = 1/2$.

c) Suppose that $\mathcal{H}$ is a family of 4-wise independent hash functions. Prove that it is a sketching family with error $\epsilon = 2/3$. In this part you can use the Paley-Zygmund inequality:

Theorem 2.1 (Paley-Zygmund Inequality)**.** *If $Z \geq 0$ is a R.V. and $0 \leq \alpha \leq 1$, then*

$$\mathbb{P}\left[Z > \alpha \mathbb{E}\left[Z\right]\right] \geq (1 - \alpha)^2 \frac{\mathbb{E}\left[Z\right]^2}{\mathbb{E}\left[Z^2\right]}.$$

2) In the maximum cut problem we are given a graph $G = (V, E)$ we want to find $\max_{(S,\overline{S})} |E(S, \overline{S})|$. Unlike the mincut problem, this problem is NP-complete so we don't expect to find the optimum solution efficiently. Instead, we want to find an approximate solution.

   a) Show that the optimum solution of this problem is always at most $|E|$. Can you construct a graph G with $m = |E|$ edges (for any $m > 0$) such that the optimum of maxcut is equal to $|E|$?

   b) Design a randomized (polynomial-time) algorithm that outputs a (random) cut $(S, \overline{S})$ such that
   $$\mathbb{E}\left[|E(S, \overline{S})|\right] \geq |E|/2.$$
   So, such an algorithm gives a 2-approximation for max-cut (in expectation).

   c) Design a randomize (polynomial-time) algorithm that uses only $O(\log n)$-many random bits and returns a cut $(S, \overline{S})$ such that
   $$\mathbb{E}\left[|E(S, \overline{S})|\right] \geq |E|/2.$$

   d) Design a *deterministic* (polynomial-time) algorithm that outputs a cut $(S, \overline{S})$ such that
   $$|E(S, \overline{S})| \geq |E|/2.$$

3) In this problem we design an LSH for points in $\mathbb{R}^d$, with the ℓ_1 distance, i.e.
$$d(p, q) = \sum_i |p_i - q_i|.$$

   a) Let a, b be arbitrary real numbers. Fix $w > 0$ and let $s \in [0, w)$ chosen uniformly at random. Show that
   $$\mathbb{P}\left[\left\lfloor \frac{a-s}{w} \right\rfloor = \left\lfloor \frac{b-s}{w} \right\rfloor\right] = \max\left\{0, 1 - \frac{|a-b|}{w}\right\}.$$
   Recall that for any real number c, $\lfloor c \rfloor$ is the largest integer which is at most c.
   Hint: Start with the case where $a = 0$.

   b) Define a class of hash functions as follows: Fix w larger than diameter of the space. Each hash function is defined via a choice of d independently selected random real numbers $s_1, s_2, \ldots, s_d$, each uniform in $[0, w)$. The hash function associated with this random set of choices is
   $$h(x_1, \ldots, x_d) = \left(\left\lfloor \frac{x_1 - s_1}{w} \right\rfloor, \left\lfloor \frac{x_2 - s_2}{w} \right\rfloor, \ldots, \left\lfloor \frac{x_d - s_d}{w} \right\rfloor\right).$$
   Let $\alpha_i = |p_i - q_i|$. What is the probability that $h(p) = h(q)$ in terms of the α_i values? For what values of p_1 and p_2 is this family of functions $(r, c \cdot r, p_1, p_2)$-sensitive? Do your calculations assuming that $1 - x$ is well approximated by e^{-x}.

4) Let $u, v \in \mathbb{R}^d$ and $g \in \mathbb{R}^d$ be a random Gaussian vector, i.e., for each $1 \leq i \leq d$, $g_i \sim \mathcal{N}(0, 1)$.

   a) What is the expected value of $\langle g, u \rangle$?

   b) What is the expected value of $\langle g, u \rangle \cdot \langle g, v \rangle$?

   c) What is the expected value of $|\langle g, u \rangle|$? You can use that p.d.f. of a $\mathcal{N}(0, 1)$ $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

   d) Consider the following hash function: $h_g(u) = \text{sgn}(\langle g, u \rangle)$, where sgn is the sign function, i.e.,
   $$\text{sgn}(a) = \begin{cases} +1 & \text{if } a \geq 0 \\ -1 & \text{otherwise.} \end{cases}$$
   Show that for a random Gaussian vector g and any two vectors u, v, $\mathbb{P}[h_g(u) = h_g(v)] = 1 - \frac{\theta(u,v)}{\pi}$ where $\theta(p, q)$ is the angle between the vector of p and q.

e) Let $P \subseteq \mathbb{R}^d$ and consider the following distance function: $\text{dist}(p, q) = \frac{\theta(p,q)}{\pi}$. For what values of p_1 and p_2 is this family of functions $(r, c \cdot r, p_1, p_2)$-sensitive?

5) **Extra Credit:** Say we have a plane with n seats and we have a sequence of n passengers $1, 2, \ldots, n$ who are going to board the plane in this order and suppose passenger i is supposed to sit at seat i. Say when 1 comes he chooses to sit at some arbitrary seat different from his own sit, 1. From now on, when passenger i boards, if her seat i is available she sits at i, otherwise she chooses sits at a uniformly random seat that is still available. What is the probability that passenger n sits at her seat n?