1 Bounding global sensitivity

1. Let $f : X^n \to \mathbb{R}^d$. Let A be a (deterministic) algorithm with the following properties:
   - On input $x \in X^n$, A operates on a subsample $\tilde{x}$ of x where each entry of x appears in $\tilde{x}$ with probability at most $\alpha < 1$, and
   - $\Pr[\|A(\tilde{x}) - f(x)\|_1 \leq \sigma] > \frac{1+\alpha}{2}$.

   Show that $GS_f \leq 2\sigma$.

   Hint: consider two neighboring databases x, x' that differ on location i and let $\mathcal{E}$ be the event that entry i is not selected to be in the subsample. Note that conditioned on $\mathcal{E}$, subsamples of x and x' are identically distributed. Write $\Pr[\|A(\tilde{x}) - f(x)\|_1 > \sigma]$ as $\Pr[\mathcal{E}] \cdot \Pr[\|A(\tilde{x}) - f(x)\|_1 > \sigma | \mathcal{E}] + \Pr[\neg\mathcal{E}] \cdot \Pr[\|A(\tilde{x}) - f(x)\|_1 > \sigma | \neg\mathcal{E}]$.

2. Let median : $[0,1]^n \to [0,1]$ be the function that on input $x \in [0,1]^n$ returns the $\lceil n/2 \rceil$-th element in a sorting of x. What is GS_{median}?

2 Group privacy

Prove that differential privacy provides protection not only to individuals but also to groups of size t.

1. **Pure privacy:** Let $M : X^n \to R$ be $(\epsilon, 0)$-differentially private. Show that for all $x, x' \in X^n$ that differ on t elements and for all $T \subset R$

$$\Pr[M(x) \in T] \leq e^{t\epsilon} \cdot \Pr[M(x') \in T].$$

2. **Approximate privacy:** Let $M : X^n \to R$ be (ϵ, δ)-differentially private. Show that for all $x, x' \in X^n$ that differ on t elements and for all $T \subset R$

$$\Pr[M(x) \in T] \leq e^{t\epsilon} \cdot \Pr[M(x') \in T] + t \cdot e^{t\epsilon} \cdot \delta.$$

In both parts, the probability is over the randomness of the mechanism M. We say that x, x' differ on t elements if $|\{i : x_i \neq x'_i\}| = t$. (In particular, x, x' that differ on one position are neighboring.)

3 Noise magnitude for count queries, Laplace mechanism

Let $x \in \{0,1\}^n$ and consider the function $f(x) = \sum_{i=1}^{n} x_i$.

1. We saw that the randomized algorithm $A(x) = f(x)+Y$ where $Y \sim \text{Lap}(1/\epsilon)$ is ϵ-differentially private. Show that for all $x \in \{0,1\}^n$,

$$\Pr\left[|A(x) - f(x)| \geq \frac{\ln(\frac{1}{\delta})}{\epsilon}\right] \leq \delta.$$

2. Prove that for any $(\epsilon, 0)$-differentially private (approximation) algorithm A there exists $x \in \{0,1\}^n$ for which

$$\Pr\left[|A(x) - f(x)| \geq \frac{\ln(\frac{1-\delta}{\delta})}{2\epsilon}\right] \geq \delta.$$

In both parts of the question, the probability is taken over the randomness of the approximation algorithm, A.

Hint for part 2: consider instances x, x' that are at Hamming distance $\frac{\ln(\frac{1-\delta}{\delta})}{\epsilon}$ apart. Assume that $\Pr\left[|A(x) - f(x)| > \frac{\ln(\frac{1-\delta}{\delta})}{2\epsilon}\right] \leq \delta$ and conclude that the inequality in part 2 holds for x'.

4 Randomization

Show that a non-trivial differentially private algorithm has to be randomized. More specifically, that if a deterministic algorithm $\mathcal{A}$ does not output the same answer on all inputs, it is *not* differentially private.

5 Differentially Private Elections

A function *majority* on 0/1 inputs is defined as follows: $f_{maj}(x_1, \ldots, x_n)$ is 1 when $\geq n/2$ arguments are 1, and 0 otherwise. Give an ε-differentially private algorithm with the following property: if the input contains $\geq n/2 + k$ occurrences of bit b then your algorithm should output b with probability at least $1 - e^{\varepsilon \cdot k/4}$. (*Hint:* Use the global sensitivity framework.)

6 Median-finding using sum queries

Given a set X of n real numbers $x_1, ..., x_n$ in $[0,1]$, the rank of a value y is the number of indices i such that $x_i \leq y$. We say y is a median of X if it has rank $\lceil n/2 \rceil$. Give a differentially private algorithm which takes X as input and approximates the median in the following sense: after asking t questions with global sensitivity 1, with probability at least 2/3, the algorithm should output an interval of width 2^{-t} that contains a value with rank $\frac{n}{2} \pm \frac{t\log(t)}{\epsilon}$.

You may want to use (and first prove!) the following lemma:

Lemma 1 *Let $Z_1, Z_2, \ldots, Z_t$ be a collection of independent Laplace random variables with scale parameter 1, and let $M = \max(Z_1, ..., Z_n)$. Then*

$$(\forall x > 0)\ \Pr(M > x) \leq \tfrac{1}{2} t \exp(-x) \qquad \textit{and} \qquad \mathbb{E}(M) \leq \ln t\,.$$

7 Balanced cut

Given an undirected graph G with n vertices, a *balanced cut* is a partition of G into two disjoint sets of size at least $\lfloor n/2 \rfloor$ each. The *weight* of a cut is the number of edges that cross between the two sets. Let $OPT(G)$ denote the weight of the lightest balanced cut.

1. Give an (inefficient time $2^{O(n)}$) randomized algorithm A that outputs a cut with expected weight $OPT(G) + O(n/\varepsilon)$. Your algorithm should satisfy "edge privacy", that is, for any two graphs G and G' that differ in a single edge, and for every event E, $\Pr(A(G) \in E) \leq e^{\varepsilon} \Pr(A(G') \in E)$.

2. Suppose we create a graph G on n vertices by adding an edge between every pair of vertices independently with probability $1/2$. Show that the expected size of the $OPT(G)$ is $\Omega(n^2)$.