Data Privacy Homework 1

傅申　PB20000051

1. (15′) Laplace Mechanism

(a) **(5′)** Given the function $f(x) = \frac{1}{6}\sum_{i=1}^{6} x_i$, where $x_i \in \{1, 2, ..., 10\}$ for $i \in \{1, 2, ..., 6\}$, compute the global sensitivity and local sensitivity when $x = \{3, 5, 4, 5, 6, 7\}$.

(b) **(10′)** Given a database x where each element is in $\{1, 2, 3, 4, 5, 6\}$, design ε-differentially private Laplace mechanisms corresponding to the following queries, where $\varepsilon = 0.1$:

1. $q_1(x) = \sum_{i=1}^{6} x_i$
2. $q_2(x) = \max_{i\in\{1,2,...,6\}} x_i$

2. (15′) Exponential mechanism

ID	sex	Chinese	Mathematics	English	Physics	Chemistry	Biology
1	Male	96	58	80	53	56	100
2	Male	60	63	77	50	59	75
3	Female	83	86	98	69	80	100
				…			
4000	Female	86	83	98	87	82	92

Table 1: Scores of students in School A

Table 1 records the scores of students in School A in the final exam. We need to help the teacher query the database while protecting the privacy of students' scores. The domain of this database is $\{\text{Male}, \text{Female}\} \times \{0, 1, 2, ..., 100\}^6$. Answer the following questions.

(a) **(5′)** What is the sensitivity of the following queries:

1. $q_1(x) = \frac{1}{4000}\sum_{\text{ID}=1}^{4000} \text{Physics}_{\text{ID}}$
2. $q_2(x) = \max_{\text{ID} \in\{1,2,...,4000\}} \text{Biology}_{\text{ID}}$

(b) **(10′)** Design ε-differentially privacy mechanisms corresponding to the two queries in (a), where $\varepsilon = 0.1$. (Using Laplace mechanism for q_1 and exponential mechanism for q_2)

3. (20′) Composition

Theorem 3.16. Let $\mathcal{M}_i : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_i$ be an $(\varepsilon_i, \delta_i)$-differentially private algorithm for $i \in [k]$. Then if $\mathcal{M}_{[k]}(x) = \left(\mathcal{M}_1(x), ..., \mathcal{M}_{k(x)}\right)$, then $\mathcal{M}_k$ is $\left(\sum_{i=1}^{k} \varepsilon_i, \sum_{i=1}^{k} \delta_i\right)$-differentially private.

Theorem 3.20 (Advanced Composition). For all $\varepsilon, \delta, \delta' \geq 0$, the class of (ε, δ)-differentially private mechanisms satisfies $\left(\varepsilon', k\delta + \delta'\right)$-differential privacy under k-fold adaptive composition for:

$$\varepsilon' = \varepsilon\sqrt{2k\ln\left(\frac{1}{\delta'}\right)} + k\varepsilon(\varepsilon^{\varepsilon} - 1)$$

(a) **(10′)** Given a database $x = \{x_1, x_2, ..., x_{2000}\}$ where $x_i \in \{0, 1, 2, ..., 100\}$ for each i and privacy parameters $(\varepsilon, \delta) = \left(1.25, 10^{-5}\right)$, apply the Gaussian mechanism to protect 100 calls to the query $q_1(x) = \frac{1}{2000}\sum_{i=1}^{2000} x_i$. Determine the noise variances σ^2 of the Gaussian mechanism to ensure (ε, δ)-DP based on the composition and advanced composition theorems, respectively.

(b) **(10′)** Determine the noise variances σ^2 of the Gaussian mechanism to protect 100 calls to the query $q_2(x) = \max_{i\in\{1,2,...,2000\}} x_i$ to ensure $\left(1.25, 10^{-5}\right)$-DP based on the composition and advanced composition theorems, respectively, where x is the database in (a).

4. (25′) Randomized Response for Local DP

Consider a population of n users, where the true proportion of males is denoted as π. Our objective is to gather statistics on the proportion of males, prompting a sensitive question: "Are you male?" Each user responds with either a yes or no, but due to privacy concerns, they refrain from directly disclosing their true gender. Instead, they employ a biased coin with a probability of landing heads denoted as p, and tails as $1-p$. When the coin is tossed, a truthful response is given if heads appear, while the opposite response is provided if tails come up.

(a) **(10′)** Demonstrate that the aforementioned randomized response adheres to local differential privacy and determine the corresponding privacy parameter, ε.

(b) **(15′)** Employing the perturbation method outlined above to aggregate responses from the n users yields a statistical estimate for the number of males. Assuming the count of "yes" responses is n_1, construct an unbiased estimate π for based on n, n_1, p. Calculate the variance associated with this estimate.

5. (10′) Accuracy Guarantee of DP

Consider the application of an (ε, δ)-differentially private Gaussian mechanism denoted by $\mathscr{M}$ to protect the mean estimator $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ of a d-dimensional input database x, where $x_i \in \{0, 1, \ldots, 100\}^d$ for each i. Let $\mathscr{M}(x)$ represent the output of this Gaussian mechanism. Utilize both the tail bound and the union bound to derive the L_∞-norm error bound of $\mathscr{M}$, denoted by $\|\mathscr{M}(x) - \bar{x}\|_\infty$, ensuring a probability of at least $1-\beta$. Specifically, solve for the bound $\mathscr{B}$ such that

$$\Pr[\|\mathscr{M}(x) - \bar{x}\|_\infty \leq \mathscr{B}] \geq 1 - \beta$$

Hint: Refer to Zhihu link for descriptions of statistical inequalities.

6. (15′) Personalized Differential Privacy

Consider an n-element dataset D where the i-th element is owned by a user $i \in [n]$, where $[n] = \{1, 2, \ldots, n\}$ and the privacy requirement of user i is ε_i-DP. A randomized mechanism $\mathscr{M}$ satisfies $\{\varepsilon_i\}_{i\in[n]}$-personalized differential privacy (or $\{\varepsilon_i\}_{i\in[n]}$-PDP) if, for every pair of neighboring datasets D, D' differing at the j-th element for an arbitrary $j \in [n]$, and for all sets S of possible outputs,

$$\Pr[\mathscr{M}(D) \in S] \leq \exp(\varepsilon_j) \Pr\left[\mathscr{M}\left(D'\right) \in S\right].$$

(a) **(5′)** Prove the composition theorem of PDP: if a mechanism is $\left\{\varepsilon_i^{(1)}\right\}_{i\in[n]}$-PDP and another is $\left\{\varepsilon_i^{(2)}\right\}_{i\in[n]}$-PDP, then publishing the result of both is $\left\{\varepsilon_i^{(1)} + \varepsilon_i^{(2)}\right\}_{i\in[n]}$-PDP.

(b) **(10′)** Given a dataset D and a privacy requirement set $\{\varepsilon_i\}_{i\in[n]}$, the *Sample mechanism* works as follows:

1) We pick an arbitrary threshold value $t > 0$;
2) We sample a subset $D_S \subset D$ where the probability that the i-th element of D is contained in D_S equals $\frac{\exp(\varepsilon_i)-1}{\exp(t)-1}$ if $\varepsilon_i < t$ and 1 otherwise;
3) We output $\mathscr{M}(D_S)$, where $\mathscr{M}$ is a t-differentially private mechanism.

Prove that the Sample mechanism with any $t > 0$ is $\{\varepsilon_i\}(i \in [n])$-PDP.
Hint: Use the Bayes formula.