DataPrivacy—hw2

Terence Wang

2023/12/08

Contents

1 Q1

1.1 a

$global\ sensitivity = \max\limits_{H(D,D')=1} ||f(DB) - f(DB')||_1$, thus $global\ sensitivity = \frac{1}{6} \times (10-1) = 1.5$
$local\ sensitivity(D) = \max\limits_{D' \in N(D)} |f(D) - f(D')| = \frac{1}{6} \times (10-3) = \frac{7}{6}$

1.2 b

1.2.1

$q_1(x) = \sum\limits_{i=1}^{6} x_i$
So we can get $\Delta q_1 = 6 - 1 = 5$. Thus $M_L(x, q_1(\cdot), \epsilon = 0.1) = q_1(x) + (Y_1, \cdots, Y_6)$, where Y_i are i.i.d. random variables drawn from $Lap(\frac{\Delta q_1}{\epsilon}) = Lap(50)$.

1.2.2

$q_2(x) = \max\limits_{i \in \{1,2,\cdots,6\}} x_i$
So we can get $\Delta q_2 = 6 - 1 = 5$. Thus $M_L(x, q_2(\cdot), \epsilon = 0.1) = q_2(x) + (Y_1, \cdots, Y_6)$, where Y_i are i.i.d. random variables drawn from $Lap(\frac{\Delta q_2}{\epsilon}) = Lap(50)$.

2 Q2

2.1 a

2.1.1

$q_1(x) = \frac{1}{4000} \sum\limits_{ID=1}^{4000} Physics_{ID}$
$sensitivity = \frac{1}{4000}(100 - 0) = 0.025$

2.1.2

$q_2(x) = \max_{ID \in \{1,2,\cdots,4000\}} Biology_{ID}$
$sensitivity = 100 - 0 = 100$

2.2 b

2.2.1

$\Delta q_1 = 0.025$, thus $M_L(x, q_1(\cdot), \epsilon = 0.1) = q_1(x) + (Y_1, \cdots, Y_{4000})$, where Y_i are i.i.d. random variables drawn from $Lap(\frac{\Delta q_1}{\epsilon}) = Lap(0.25)$.

2.2.2

$\Delta u = 100$, $\epsilon = 0.1$, thus we will output with the probability $\propto$ $exp(\frac{\epsilon q_2(x)}{2\Delta u}) = exp(\frac{q_2(x)}{2000})$

3 Q3

3.1 a

3.1.1

$M_{[100]}(x)$ satisfies $(\sum_{i=1}^{100} \epsilon_i, \sum_{i=1}^{100} \delta_i) - DP = (100\epsilon_0, 100\delta_0) - DP$
Thus, $\epsilon_0 = 1.25 \times 10^{-2}$, $\delta_0 = 1 \times 10^{-7}$. $\Delta q_1 = \frac{100}{2000} = 0.05$, therefore $\sigma^2 = \frac{2\ln(\frac{1.25}{\delta_0}) \times (\Delta q_1)^2}{(\epsilon_0)^2} = 522.92$

3.1.2

$\epsilon' = 1.25$, $100 \times \delta + \delta' = 10^{-5}$. $\delta' = \delta \to \delta' = \frac{10^{-5}}{101} = 9.9 \times 10^{-8}$
According to $\epsilon' = \sqrt{2k\ln(1/\delta')}\epsilon + k\epsilon(e^{\epsilon} - 1)$, we can get $1.25 = \sqrt{2 \times 100 \times \ln(\frac{1}{9.9 \times 10^{-8}})} \times \epsilon + 100 \times \epsilon(e^{\epsilon} - 1)$. Therefore, $\epsilon = 0.02121$. $\Delta q_1 = \frac{100}{2000} = 0.05$, thus $\sigma^2 = \frac{2\ln(\frac{1.25}{\delta}) \times (\Delta q_1)^2}{(\epsilon)^2} = 181.74$

3.2 b

3.2.1

$M_{[100]}(x)$ satisfies $(\sum_{i=1}^{100}\epsilon_i, \sum_{i=1}^{100}\delta_i) - DP = (100\epsilon_0, 100\delta_0) - DP$
Thus, $\epsilon_0 = 1.25 \times 10^{-2}$, $\delta_0 = 1 \times 10^{-7}$. $\Delta q_2 = 100 - 0 = 100$, therefore $\sigma^2 = \frac{2\ln(\frac{1.25}{\delta_0})\times(\Delta q_2)^2}{(\epsilon_0)^2} = 2091678618$

3.2.2

$\epsilon' = 1.25$, $100 \times \delta + \delta' = 10^{-5}$. $\delta' = \delta \to \delta' = \frac{10^{-5}}{101} = 9.9 \times 10^{-8}$
According to $\epsilon' = \sqrt{2k\ln(1/\delta')}\epsilon + k\epsilon(e^\epsilon - 1)$, we can get $1.25 = \sqrt{2 \times 100 \times \ln(\frac{1}{9.9\times 10^{-8}})} \times \epsilon + 100 \times \epsilon(e^\epsilon - 1)$. Therefore, $\epsilon = 0.02121$. $\Delta q_2 = 100 - 0 = 100$, thus $\sigma^2 = \frac{2\ln(\frac{1.25}{\delta})\times(\Delta q_2)^2}{(\epsilon)^2} = 726943516.4$

4 Q4

4.1 a

$\frac{Pr[f(t)=t^\star]}{Pr[f(t')=t^\star]} \leq \frac{p}{1-p}$, let $\epsilon = \ln\frac{p}{1-p}$, then we get $Pr[f(t) = t^\star] \leq e^\epsilon Pr[f(t') = t^\star]$. So the aforementioned randomized response adheres to local differential privacy, and $\epsilon = \ln\frac{p}{1-p}$.

4.2 b

$P(X_i = yes) = \pi p + (1-\pi)(1-p)$, $P(X_i = no) = (1-\pi)p + \pi(1-p)$
Construct the likelihood function $L = \prod^{n_1}[\pi p + (1-p)(1-\pi)] \prod^{n-n_1}[(1-\pi)p + \pi(1-p)] = [\pi p + (1-p)(1-\pi)]^{n_1}[(1-\pi)p + \pi(1-p)]^{n-n_1}$
Take the logarithm: $\ln(L) = n_1\ln[\pi p + (1-p)(1-\pi)] + (n-n_1)\ln[(1-\pi)p + \pi(1-p)]$
Take the derivative of the variable π and set the derivative to 0: $\frac{\partial\ln(L)}{\partial\pi} = \frac{n_1}{\pi p + (1-p)(1-\pi)} \times (p - (1-p)) - \frac{n-n_1}{(1-\pi)p + \pi(1-p)} \times (p - (1-p)) = 0$
Therefore, we can get $\hat{\pi} = \frac{p-1}{2p-1} + \frac{n_1}{(2p-1)n}$

$E(\hat{\pi}) = \frac{1}{2p-1}[p-1+\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{2p-1}[p-1+\frac{1}{n}\cdot n \cdot Pr[X_i = yes]] = \frac{1}{2p-1}[p-1+\pi p+(1-\pi)(1-p)] = \pi$. Thus $\hat{\pi}$ is an unbiased estimator of π.
$Var(\hat{\pi}) = Var(\frac{n_1}{(2p-1)n}) = \frac{1}{(2p-1)^2n^2}Var(n_1) = \frac{(1+2\pi p-\pi-p)(\pi+p-2\pi p)}{(2p-1)^2 n}$

5 Q5

$B = \sqrt{2\ln(1.25/\delta)\ln(d/\beta)}\frac{\Delta_2(f)}{\epsilon}$
According to the theorem: Figure 1

Theorem 2.2 (Gaussian Mechanism, Dwork & Roth (2014)). *Let $\epsilon > 0$ and $\delta > 0$. For*

9

any algorithm f mapping a data set $\mathcal{D}$ to $\mathbb{R}^d$, the Gaussian Mechanism $A(\cdot)$ defined as

$$A(\mathcal{D}) = f(\mathcal{D}) + (u_1, \ldots, u_d)^\top, \tag{8}$$

where $u_1, \ldots, u_d \overset{i.i.d.}{\sim} \mathcal{N}(0, 2\ln(1.25/\delta)(\Delta_2(f)/\epsilon)^2)$, is (ϵ, δ)-DP.

Figure 1: theorem

We can conclude that $M(x) - \overline{x}$ is $N(0, 2\ln(1.25/\delta)(\Delta_2(f)/\epsilon)^2)$
$Pr[||M(x) - \overline{x}||_\infty \le B] \ge 1-\beta$ is equal to $Pr[||M(x) - \overline{x}||_\infty > B] < \beta$.
This is equal to $Pr[\max_{i\in[d]} |M(x) - \overline{x}| > B] < \beta$.

Use **union bound**, if we can get $d \cdot Pr[|M(x) - \overline{x}| > B] < \beta$, then we have $Pr[\max_{i\in[d]} |M(x) - \overline{x}| > B] \le d \cdot Pr[|M(x) - \overline{x}| > B] < \beta$

Use **Chernoff bound**:$P(X-\mu \ge a) \le e^{-\frac{a^2}{2\sigma^2}}$, we can get $Pr[M(x) - \overline{x} > B] \le e^{-\frac{B^2}{2\sigma^2}}$. So $Pr[|M(x) - \overline{x}| > B] \le e^{-\frac{B^2}{\sigma^2}}$.
Let $\frac{\beta}{d} = e^{-\frac{B^2}{\sigma^2}}$. Therefore, we can get $B = \sqrt{2\ln(1.25/\delta)\ln(d/\beta)}\frac{\Delta_2(f)}{\epsilon}$
In this question, $\Delta_2(f) = \frac{100\sqrt{d}}{n}$, so $B = \sqrt{2\ln(1.25/\delta)\ln(d/\beta)}\frac{100\sqrt{d}}{\epsilon n}$

6 Q6

6.1 a

According to the definition of $\{\epsilon_i\}_{i\in[n]} - PDP$: $\frac{Pr[M_1(D)\in S_1]}{Pr[M_1(D')\in S_1]} \leq e^{\epsilon_i^{(1)}}$, $\frac{Pr[M_2(D)\in S_2]}{Pr[M_2(D')\in S_2]} \leq e^{\epsilon_i^{(2)}}$. Therefore, $\frac{Pr[M_{1,2}(D)\in S_1\times S_2]}{Pr[M_{1,2}(D')\in S_1\times S_2]} = \frac{Pr[M_1(D)\in S_1]\times Pr[M_2(D)\in S_2]}{Pr[M_1(D')\in S_1]\times Pr[M_2(D')\in S_2]} \leq e^{\epsilon_i^{(1)}+\epsilon_i^{(2)}}$.
Thus, publishing the result of both is $\{\epsilon_i^{(1)} + \epsilon_i^{(2)}\}_{i\in[n]}$-PDP

6.2 b

$\pi_i = \begin{cases} \frac{e^{\epsilon_i}-1}{e^t-1} & \epsilon_i\text{<t} \\ 1 & \text{otherwise} \end{cases}$
Let $D_{S_{-i}}(D_{S_{+i}})$ denote the dataset resulting from removing(adding) the i-th element from D_S.
Let DP denote any t-differentially private mechanism.
Let RS denote the procedure that samples each element.
So the Sample mechanism can be defined as $M(D_S) = DP(RS(D_S))$
We want to prove $\frac{Pr[M(D_S)\in S]}{Pr[M(D_{S_{-i}})\in S]} \leq e^{\epsilon_i}$
$Pr[M(D_S) \in S] = \sum\limits_{Z\subset D_{S_{-i}}} (\pi_i \cdot Pr[RS(D_{S_{-i}}) = Z] \cdot Pr[DP(D_{S_{+i}}) \in S]) + ((1-\pi_i) \cdot Pr[M(D_{S_{-i}}) \in S])$
Since DP is t-differentially private, we can get $Pr[DP(D_{S_{+i}}) \in S] \leq e^t \cdot Pr[DP(D_{S_i}) \in S]$
Therefore, $Pr[M(D_S) \in S] \leq \sum\limits_{Z\subset D_{S_{-i}}} (\pi_i \cdot Pr[RS(D_{S_{-i}}) = Z]\cdot e^t \cdot Pr[DP(D_{S_i}) \in S]) + ((1-\pi_i) \cdot Pr[M(D_{S_{-i}}) \in S]) = \pi_i(e^t \cdot Pr[M(D_{S_{-i}}) \in S]) + (1-\pi_i)Pr[M(D_{S_{-i}}) \in S] = (1-\pi_i+\pi_i e^t)Pr[M(D_{S_{-i}}) \in S]$
If $\epsilon_i \geq t$**:**
We can get $\pi_i = 1$, so $Pr[M(D_S) \in S] = (1-1+e^t)Pr[M(D_{S_{-i}}) \in S] = e^t Pr[M(D_{S_{-i}}) \in S] \leq e^{\epsilon_i} Pr[M(D_{S_{-i}}) \in S]$
If $\epsilon_i < t$**:**
We can get $\pi_i = \frac{e^{\epsilon_i}-1}{e^t-1}$, so $Pr[M(D_S) \in S] \leq \frac{e^{\epsilon_i}-1}{e^t-1}(e^t Pr[M(D_{S_{-i}}) \in S]) + (1-\frac{e^{\epsilon_i}-1}{e^t-1})Pr[M(D_{S_{-i}}) \in S] = e^{\epsilon_i}Pr[M(D_{S_{-i}}) \in S]$
Thus, we prove that the Sample mechanism is $\{\epsilon_i\}_{i\in[n]}$-PDP.