COM SCI 260B HW 4 Solution

Ashish Kumar Singh (UID:105479019)

May 22, 2022

Problem 1. ϵ-differential privacy

Solution 1. We sample Z uniformly from the interval $[-\frac{C}{\epsilon}, \frac{C}{\epsilon}]$. Its pdf can be defined as:

$$pdf(z) = \begin{cases} \frac{\epsilon}{2C} & \text{for } z \in [-\frac{C}{\epsilon}, \frac{C}{\epsilon}] \\ 0 & \text{otherwise} \end{cases}$$

To check sensitivity, we compare probability of neighbourhood dataset X and X',

$$\frac{Pr[M(X) = a]}{Pr[M(X') = a]} = \frac{Pr[f(X) + Z = a]}{Pr[f(X') + Z = a]}$$

$$\frac{Pr[M(X) = a]}{Pr[M(X') = a]} = \frac{Pr[Z = a - f(X)]}{Pr[Z = a - f(X')]}$$

Using pdf we can get only three possible values for above equation, 0,1 or∞

$$\frac{Pr[M(X) = a]}{Pr[M(X') = a]} = \begin{cases} 0 \\ 1 \\ \infty \end{cases}$$

which is independent of ϵ, the issue is when $a - f(X)$ lies in range $[-\frac{C}{\epsilon}, \frac{C}{\epsilon}]$ but $a - f(X')$ lies outside. hence the mechanism is **not ϵ-deferentially private**.

Problem 2. Exponential Mechanism guarantee

Solution 2. For 2ϵ-deferentially private exponential mechanism we know that,

$$Pr\big[M_E(X,u,R)=\gamma\big] \propto exp\left(\frac{\epsilon u(X,\gamma)}{\Delta u}\right)$$

To find probability of bounded utility, we can sum all valid γ, as follows:

$$Pr\big[u(M_E(X,u,R)) \leq C\big] \propto \sum_{\gamma \in R: u(X,\gamma) \leq C} exp\left(\frac{\epsilon u(X,\gamma)}{\Delta u}\right)$$

On removing proportionality,

$$Pr\big[u(M_E(X,u,R)) \leq C\big] = \frac{\sum_{\gamma \in R: u(X,\gamma) \leq C} exp\left(\frac{\epsilon u(X,\gamma)}{\Delta u}\right)}{\sum_{\gamma \in R} exp\left(\frac{\epsilon u(X,\gamma)}{\Delta u}\right)}$$

We can upper bound the numerator by using total items $|R|$ and lower bound the denominator as there exist atleast one γ which gives optimum utility $OPT_u(X)$

$$Pr\big[u(M_E(X,u,R)) \leq C\big] \leq \frac{|R| exp\left(\frac{\epsilon C}{\Delta u}\right)}{exp\left(\frac{\epsilon OPT_u(X)}{\Delta u}\right)}$$

$$Pr\big[u(M_E(X,u,R)) \leq C\big] \leq |R| exp\left(\frac{\epsilon(C - OPT_u(X))}{\Delta u}\right)$$

Using $C = OPT_u(X) - \frac{\Delta u}{\epsilon}(ln|R| + t)$ in above equation we get,

$$Pr\big[u(M_E(X,u,R)) \leq OPT_u(X) - \tfrac{\Delta u}{\epsilon}(ln|R|+t)\big] \leq |R| exp\big(-ln|R| - t\big)$$

$$Pr\big[u(M_E(X,u,R)) \leq OPT_u(X) - \tfrac{\Delta u}{\epsilon}(ln|R|+t)\big] \leq |R| e^{-ln|R|} e^{-t}$$

$$Pr\big[u(M_E(X,u,R)) \leq OPT_u(X) - \tfrac{\Delta u}{\epsilon}(ln|R|+t)\big] \leq e^{-t}$$

which proves the required guarantee.

Problem 3. Median income

Solution 3a. Since each person income is integer in range $[0, N]$, If n is odd then changing one person's income can change the median atmost by N, if n is even the the same would change atmost by $N/2$, hence the sensitivity is atmost N.

$$S_1(f) = max_{X,X'}||f(X) - f(X')||_1$$

$$S_1(f) = N$$

Solution 3b. Since the sensitivity from above is N, we have to add Laplacian noise (N/ϵ) to achieve ϵ-differential privacy. But this is too much noise and the error with actual median will be very high. The added noise is directly proportional to N, hence achieving ϵ-differential privacy through Laplacian mechanism is not scalable for median reporting.

Solution 3c. We can define a utility function to give the negative of the minimum L0 norm of difference with X when median is γ,

$$u(X, \gamma) = -min_{Y:median(Y)=\gamma}||X - Y||_0$$

Sensitivity can be computed as follows,

$$S_1(f) = max_{X,X'}|u(X, \gamma) - u(X', \gamma)|$$

$$S_1(f) = |min_{Z:median(Z)=\gamma}||X' - Z||_0 - min_{Y:median(Y)=\gamma}||X - Y||_0|$$

We know that $||X - X'||_0 = 1$, as they are neighbouring dataset.
Lets assume $min_{Y:median(Y)=\gamma}||X - Y||_0 = c$. Using inequality for L0 norm, we can write:

$$||X' - Y||_0 \leq ||X' - X||_0 + ||X - Y||_0$$

$$||X' - Y||_0 \leq 1 + c$$

There exists a Y with above inequality, then Z (argmin) should also satisfy it,

$$||X' - Z||_0 \leq 1 + c$$

Using these in the equation for sensitivity, we get,

$$S_1(f) \leq |1 + c - c|$$

$$S_1(f) \leq 1$$

Hence the sensitivity is atmost 1 which is independent of N, n.

Solution 3d. We can use the same utility as in previous question but with 90 percentile condition.

$$u(X, \gamma) = -min_{Y:90percentile(Y)=\gamma}||X - Y||_0$$

Sensitivity can be computed as follows,

$$S_1(f) = max_{X,X'}|u(X, \gamma) - u(X', \gamma)|$$

$$S_1(f) = |min_{Z:90percentile(Z)=\gamma}||X' - Z||_0 - min_{Y:90percentile(Y)=\gamma}||X - Y||_0|$$

We know that $||X - X'||_0 = 1$, as they are neighbouring dataset.
Lets assume $min_{Y:90percentile(Y)=\gamma}||X - Y||_0 = c$. Using inequality for L0 norm, we can write:

$$||X' - Y||_0 \leq ||X' - X||_0 + ||X - Y||_0$$

$$||X' - Y||_0 \leq 1 + c$$

There exists a Y with above inequality, then Z (argmin) should also satisfy it,

$$||X' - Z||_0 \leq 1 + c$$

Using these in the equation for sensitivity, we get,

$$S_1(f) \leq |1 + c - c|$$

$$S_1(f) \leq 1$$

Hence the sensitivity is atmost 1 which is independent of N, n.