CS325: Private Data Analysis **Fall 2022**

Lecture 10: Beyond Global Sensitivity

Lecturer: Di Wang *Scribes: Di Wang*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

So far we have focused on adding noise to a statistic calibrated to its global sensitivity. That is, we consider the maximum sensitivity over **all possible neighbouring datasets** :

$$\Delta_2(f) = \max_{D\sim D'} \|f(D) - f(D')\|_2,$$

$$\Delta_1(f) = \max_{D\sim D'} \|f(D) - f(D')\|_1,$$

However, these global sensitivities sometimes could be very large: For example we let $f(D)$ to be the median of the dataset where each record $x_i \in [0, \Lambda]$. We can easily see that its global sensitivity is Λ. The reason is as the following: Suppose $n = 2m + 1$, in D $x_1 = \cdots = x_{m+1} = 0, x_{m+2} = \cdots = x_n = \Lambda$ and in D', $x_1 = \cdots = x_m = 0, x'_{m+1} = \cdots = x_n = \Lambda$ we can see $|f(D) - f(D')| = \Lambda$. To address the large sensitivity issue, one possible way is instead of adding noise related to the Global sensitivity, which is the worst case, here we add noise related to the Local sensitivity, which depends on the input dataset:

Definition 10.1 (Local Sensitivity) *The local sensitivity of a function $f : \mathcal{X}^n \mapsto \mathbb{R}$ on a dataset D is defined as*

$$\Delta_{LS}(f)(D) = \max_{D\sim D} |f(D) - f(D')|.$$

With this definition in mind, the natural approach would be to add Laplace noise calibrated to the local sensitivity, rather than the global sensitivity as we have been doing. However, the issue is that the magnitude of the noise may reveal information about the dataset. Lets see this illustrated through an example. Consider the function f which computes the distance between the closest two points in a dataset, where the domain is the reals. Let $D_1 = \{0, 0, 0\}$ then the local sensitivity of the query on D_1 is 0. Consider another dataset where $D_2 = \{0, 0, 1000\}$. Then local sensitivity of the query on D_1 is 1000. Thus, when the adversary can get the results, he can easily identify whether the input data is D_1 or D_2.

Thus, the question is, how to deal with the case where the global sensitivity is large or even unbounded? In this lecture, we will mention two approaches.

10.1 Propose-Test-Release Method

The issue with the previous example was that, while the local sensitivity on the dataset X was low, the local sensitivity on a nearby dataset was high. Instead, we can (privately) check the distance to the nearest dataset with high sensitivity. If it is distant, then it is will be private to add a small amount of noise. If it is close to a dataset with high sensitivity, then the algorithm can give up it may return $\perp$, or it may noise according to the global sensitivity, if desired. The name of this approach is Propose-Test-Release [1]. It consists of the following three steps:

1. Propose an upper bound for the local sensitivity.
2. Privately test whether this is a valid upper bound
3. Release the value (privately) if this is the case.

In slightly more detail,

1. Propose a bound β on the local sensitivity
2. Compute the distance from D, to the nearest dataset D' such as $\Delta_{LS}(f)(D') \geq \beta$, name it γ. Distance is measured in terms of how many points must be changed to get from D to D'.
3. Compute $\hat{\gamma} = \gamma + \text{Lap}(\frac{1}{\epsilon})$.
4. If $\hat{\gamma} \leq \frac{\log \frac{1}{\delta}}{\epsilon}$ then return $\perp$. Otherwise return $f(D) + \text{Lap}(\frac{\beta}{\epsilon})$.

Theorem 10.2 *PTR Algorithm is* $(2\epsilon, \delta)$*-DP.*

Proof: First, we consider the probability of outputting $\perp$ under neighbouring databases. This is done deterministically depending on whether or not $\hat{\gamma}$ is above the threshold. Since neighbouring databases will result in a value of γ which differ by at most 1. the Laplace mechanism implies that $\mathbb{P}(A(D) = \perp) \leq e^{\epsilon}\mathbb{P}(A(D') = \perp)$.

Now we break the analysis into two cases in particular, we separately consider databases D depending on the local sensitivity. First, we consider the case when $\Delta_{LS}(f)(D) > \beta$. In this case, $\gamma = 0$ and using the PDF of the Laplace distribution, the probability that $\hat{\gamma} \geq \frac{\log \frac{1}{\delta}}{\epsilon}$ is at most δ. Thus, for any $T \subseteq \mathbb{R} \cap \perp$:

$$\begin{aligned} \mathbb{P}(A(D) \in T) &= \mathbb{P}(A(D) \in T \cap \{\perp\}) + \mathbb{P}(A(D) \in T \cap \mathbb{R}) \\ &\leq e^{\epsilon}\mathbb{P}(A(D) \in T \cap \{\perp\}) + \delta \end{aligned}$$

If $\Delta_{LS}(f)(D) \leq \beta$, we view this as the composition of two differentially private algorithms. The first one is the release of $\hat{\gamma}$ which is ϵ-DP. The second one simply applies the Laplace mechanism with a parameter which is a valid upper bound on the sensitivity, satisfying the desired guarantees of 2ϵ-DP. ■

While this framework is much more general, lets see a simple example of it in action, applied to histograms. Before, we were trying to estimate the count of entries in every bin for now, were just going to settle with an easier task, of just finding the most frequent element. When we were looking at histograms before, using the Laplace Mechanism, we focused on the case where the domain $\mathcal{X}$ was discrete. Indeed, as we incur an acccuracy error which decays logarithmically in $|X|$, a finite domain was necessary. However, since we are relaxing our requirement from pure differential privacy to approximate differential privacy, we will see this is no longer required. One might speculate that the weaker goal of only finding the most frequent element is also responsible for these savings. This turns out to not be the case, and we will mention that a similar stability-based approach can also be used to estimate the counts of all elements simultaneously.

We have a dataset $D \in \mathcal{X}^n$, and we wish to compute the most frequent element (the mode). Suppose the entire dataset is the same value v: we can see that this is an incredibly stable function, with a local sensitivity of 0 for all datasets at distance $\frac{n}{2}$. In particular, if we move fewer than $\frac{n}{2}$ datapoints, we will always have strictly greater than $\frac{n}{2}$ datapoints on v, resulting in the same mode. Along these lines, the distance to a dataset with a non-zero local sensitivity is very easy to compute: it is simply half the difference between the count of the most frequent and the second most frequent element. Interestingly, note that since we are trying to apply propose-test-release with a value of $\beta = 0$, we can actually release the most frequent item exactly, if it occurs frequently enough.

We can now instantiate the framework described above as follows:

1. Propose a bound 0 on the local sensitivity
2. Let γ be half the difference between the count on the most frequent and the second most frequent element in D.
3. Computer $\hat{\gamma} = \gamma + \text{Lap}(\frac{1}{\epsilon})$
4. If $\hat{\gamma} \leq \frac{\log \frac{1}{\delta}}{\epsilon}$ then return $\perp$. Otherwise return the most frequent element in D.

Note that this algorithm is (ϵ, δ)-differentially private-we save an ϵ, since we dont have to noise the function again at the end. What type of accuracy does it guarantee? If $\hat{\gamma} \geq \frac{\log 1/\delta}{\epsilon}$, then we return the most frequent element exactly. We know, by Laplace tail bounds, that the Laplace noise added to γ will be of magnitude at most $\frac{\log 1/\delta}{\epsilon}$ with probability $1 - \delta$, so this implies we need γ to be at least $2\frac{\log 1/\delta}{\epsilon}$ for this to occur. Since this is half the difference between the gap for the most and second most frequent elements, we need this gap to be at least $4\frac{\log 1/\delta}{\epsilon}$. Combining this all, we get the following theorem:

Theorem 10.3 *There exists an (ϵ, δ)-differentially private algorithm which identifies the most frequent element from an arbitrary dataset with probability at least $1 - \delta$, as long as the gap between the count of the most frequent and the second most frequent element is at least $4\frac{\log 1/\delta}{\epsilon}$.*

10.2 Smooth Sensitivity

Another powerful technique is smooth sensitivity, introduced by Nissim, Raskhodnikova, and Smith [2]. The lesson from the previous example is that the noise magnitude has to be an insensitive function. To decide on the noise magnitude we will use a smooth upper bound on the local sensitivity, namely, a function S that is an upper bound on $\Delta_{LS}(f)$ at all points and such that $\ln(S(\cdot))$ has low sensitivity.

Definition 10.4 *For $\beta > 0$, a function $S : \mathcal{X}^n \mapsto \mathbb{R}$ is a β-smooth upper bound on the local sensitivity of f if it satisfies the following requirements:*

1. *$\forall D \in \mathcal{X}^n$, $S(D) \geq \Delta_{LS}(f)(D)$.*
2. *$\forall D \sim D', S(D) \leq e^{\beta} S(D')$.*

We can see that when $S(D)$ is the global sensitivity then it is 0-smooth. When $\beta > 0$ it is a very conservative upper bound on $\Delta_{LS}(f)$. A function that is the smallest to satisfy Definition 2.1 is the smooth sensitivity of f:

Definition 10.5 *For $\beta > 0$, the β-smooth sensitivity is defined as*

$$S_{f,\beta}(D) = \max_{D' \in \mathcal{X}^n} \{\Delta_{LS}(f)(D') e^{-\beta d(D,D')}\}. \tag{10.1}$$

Theorem 10.6 *$S_{f,\beta}(D)$ is a β-smooth upper bound on the local sensitivity of f, moreover it satisfies that $S_{f,\beta}(D) \leq S(D)$ for every β-smooth upper bound.*

Proof: We can see that

$$S_{f,\beta}(D) \geq \max\{\Delta_{LS}(f)(D), \max_{D' \in \mathcal{X}^n, D' \neq D} \{\Delta_{LS}(f)(D') e^{-\beta d(D,D')}\}\} \geq \Delta_{LS}(f)(D). \tag{10.2}$$

For the second condition, we assume that $\tilde{D}$ satisfies $S_{f,\beta}(D) = \Delta_{LS}(f)(\tilde{D})\exp(-\beta d(\tilde{D}, D))$. Then for $S_{f,\beta}(D')$ with $D' \sim D$

$$\begin{aligned} S_{f,\beta}(D') &\geq \Delta_{LS}(f)(\tilde{D})\exp(-\beta d(\tilde{D}, D') \\ &\geq \Delta_{LS}(f)(\tilde{D})\exp(-\beta d(\tilde{D}, D) - \beta) \\ &\geq e^{-\beta}\Delta_{LS}(f)(\tilde{D})\exp(-\beta d(\tilde{D}, D)) = e^{-\beta} S_{f,\beta}(D). \end{aligned}$$

Next we will show that $S(D) \geq S_{f,\beta}(D)$, it is sufficient to show that $S(D) \geq \Delta_{LS}(f)(D')\exp(-\beta d(D, D')$ for all D' with $d(D', D) = k$. We will show it by induction, when $k = 0$, it is true since $S(D) \geq \Delta_{LS}(f)(D)$. Suppose it is true for k, then for $k + 1$, suppose $d(D, \tilde{D'}) = k + 1$, then there exists a dataset $\tilde{D}$ such that $d(D, \tilde{D}) = 1$ and $d(\tilde{D}, D') = k$. Thus, $S(D) \geq S(\tilde{D})\exp(-\beta)$. By using the induction $S(\tilde{D}) \geq \Delta_{LS}(f)(D')\exp(-\beta k)$. Thus we have $S(D) \geq \Delta_{LS}(f)(D')\exp(-\beta(k + 1))$. ■

Theorem 10.7 *Let $f : \mathcal{X}^n \mapsto \mathbb{R}$ be any real-valued function and let $S : \mathcal{X}^n \mapsto \mathbb{R}$ be a β-smooth upper bound on the local sensitivity of f. Then*

1. *If $\beta = \frac{\epsilon}{2(\gamma+1)}$ and $\gamma > 1$,the algorithm $A(D) = f(D) + \frac{2(\gamma+1)S(D)}{\epsilon}\eta$, where η is sampled from the distribution $h(z) \propto \frac{1}{1+|z|^\gamma}$, is ϵ-DP.*

2. *If $\beta \leq \frac{\epsilon}{2\ln(2/\delta)}$ and $\delta < 1$, the algorithm $A(D) = f(D) + \frac{2S(D)}{\epsilon}\eta$, where $\eta = Lap(1)$, is (ϵ, δ)-DP.*

Next, we will provide the smooth sensitivity of the median. Recall that $f(D)$ was defined to be the median of values in $\mathcal{X} = [0, \Lambda]$. For simplicity, assume n is odd, and the database elements are in nondecreasing order: $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq \Lambda$. We know $\Delta(f) = \Lambda$ and $\Delta_{LS}(f)(D) = \max\{x_m - x_{m-1}, x_{m+1} - x_m\}$ with $m = \frac{n+1}{2}$.For notational convenience, define $x_i = 0$ for $i \leq 0$ and $x_i = \Lambda$ for $i \geq n$.

Theorem 10.8 *The smooth sensitivity of median is*

$$S_{f,\epsilon}(D) = \max_{k=0,\cdots,n}(e^{-k\epsilon}\max_{t=0,\cdots,k+1}(x_{m+t} - x_{m+t-k-1})). \tag{10.3}$$

Before we prove the proposition, we illustrate the result with an example. Consider an instance where the points x_i are restricted to the interval $[0, 1]$ (that is, $\Lambda = 1$) and the points are evenly spaced the interval (that is, $x_i = \frac{i}{n}$ for $i = 1, \cdots, n$). In this case, $S_{f,\epsilon}(D) = \max_k e^{-\epsilon k}\frac{k+1}{n}$. The maximum occurs at $k = \frac{1}{\epsilon}$ We get $S_{f,\epsilon}(D) \leq \frac{1}{\epsilon n}$ and so the magnitude of the noise we add is $\frac{1}{\epsilon^2 n}$. For comparison, the noise magnitude for median in the global is $\frac{1}{\epsilon}$; adding noise of that magnitude essentially wipes out all information about the median since the extreme values, 0 and 1 are hard to distinguish.

References

[1] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 371–380, 2009.

[2] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.