**CS291A Introduction to Differential Privacy** **Fall 2021**

# Lecture 15: Propose-Test-Release (November 15)

*Lecturer: Yu-Xiang Wang* *Scribes: Shichang Liu*

**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.*

## 15.1 Smooth Sensitivity

### 15.1.1 Concentrated DP analysis of Smoothed Sensitivity

Recall that we are adding Laplace log-normal noise when $X$ is drawn from Laplace distribution and $Y$ is coming from a standard Gaussian distribution:

$$Z = X \cdot e^{\sigma Y}$$

in order to bound the concentrated differential privacy we need to bound the renyi divergence of two neighboring data set.

**Proposition 15.1.** *Let $f : \mathcal{X}^n \to \mathbb{R}$ and let $Z \leftarrow LLN(\sigma)$ for some $\sigma > 0$. Then, for all $s, t > 0$, the algorithm $M(x) = f(x) + \frac{1}{s} \cdot S_f^t(x) \cdot Z$ guarantees $\frac{1}{2}\varepsilon^2 - CDP$ for $\varepsilon = t/\sigma + e^{3\sigma^2/2}s$.*

We have:

$$\begin{aligned}
&D_\alpha \left(f(x) + g(x) \cdot z \| f\left(x'\right) + g\left(x'\right) z\right) \\
&= D_\alpha \left(g(x) \cdot z \| f\left(x'\right) - f(x) + g\left(x'\right) z\right) \\
&= D_\alpha \left(z \| \frac{f(x') - f(x)}{g(x)} + \frac{g\left(x'\right)}{g(x)} z\right)
\end{aligned}$$

Then, it becomes the noise on the left hand side and another noise that is dilated and translated on the right hand side.

### 15.1.2 Sketch of the proof for the Laplace-Log-Normal

Let's say for all neighboring datasets:

$$|f(x) - f\left(x'\right)| \leq g(x) \quad \text{and} \quad e^{-t} g(x) \leq g\left(x'\right) \leq e^t g(x)$$

We have algorithm:

$$M(x) = f(x) + \frac{g(x)}{s} \cdot Z \quad \text{for} \quad Z \leftarrow \mathrm{LLN}(\sigma)$$

Then, we have

$$\mathrm{D}_\alpha \left(M(x) \| M\left(x'\right)\right) = \mathrm{D}_\alpha \left(Z \| \frac{f\left(x'\right) - f(x)}{g(x)} \cdot s + \frac{g\left(x'\right)}{g(x)} \cdot Z\right).$$

$\frac{f(x')-f(x)}{g(x)}$ is bounded between $-1$ and 1, and $\frac{g(x')}{g(x)}$ is bounded between $e^{-t}$ and $e^t$. So with these two bounds we have reduced the problem to bounding the max of 2 concrete renyi divergence:

$$\mathrm{D}_\alpha\left(M(x)\|M\left(x'\right)\right) \leqslant \max \left\{D_\alpha\left(Z\|s+e^t Z\right), D_\alpha\left(s+e^t Z\|Z\right)\right\}$$

the idea of bounding these two divergence is to find some intermediate distribution such that we can decompose the renyi divergence calculation into triangular inequality, so that we can apply the dilation property and the translation property to each one of the component in the decomposition.

We will use a lemma of group privacy for CDP when each one of these Renyi divergences satisfy a linear upper bound:

**Lemma 15.2.** *Let $P, Q, R$ be probability distributions. Suppose $\mathrm{D}_\alpha(P\|R) \leq a\cdot\alpha$ and $\mathrm{D}_\alpha(R\|Q) \leq b\cdot\alpha$ for all $\alpha \in (1,\infty)$. Then, for all $\alpha \in (1,\infty)$,*

$$\mathrm{D}_\alpha(P\|Q) \leq \alpha\cdot(\sqrt{a}+\sqrt{b})^2 \leq 2\alpha\cdot(a+b).$$

Decompose what we want to bound:

$$\mathrm{D}_\alpha\left(Z\|e^t Z+s\right) = D_\alpha\left(Z-s\|e^t Z\right)$$

We can find that the lemma is not related to the order of P and Q. We will choose $Z-s$ to be P and $e^t Z$ to be Q. We can choose $R=Z$, we have to find upper bound for $D_\alpha(Z-s\|Z)$ and $D_\alpha\left(Z\|e^t Z\right)$, reducing the problem to just the translation only and dilation only. The other part is similar.

To bound the two parts separately, we need these two lemma:

**Lemma 15.3.** *Let $Z \leftarrow \mathrm{LLN}(\sigma)$ for $\sigma>0$. Let $t\in\mathbb{R}$ and $\alpha\in(1,\infty)$. Then*

$$\mathrm{D}_\alpha\left(Z\|e^t Z\right) \leq \frac{\alpha t^2}{2\sigma^2}.$$

*Proof.* $\mathrm{D}_\alpha\left(Z\|e^t Z\right) = \mathrm{D}_\alpha\left(Xe^{\sigma Y}\|Xe^{\sigma Y+t}\right) \leq \sup_x \mathrm{D}_\alpha\left(xe^{\sigma Y}\|xe^{\sigma Y+t}\right) \leq \mathrm{D}_\alpha(\sigma Y\|\sigma Y+t) \leq \frac{\alpha t^2}{2\sigma^2}$.

□

The first inequality is due to quasi-convexity of Renyi divergence. The second inequality is the closure to post processing of normal distribution. The divergence satisfies the concentrated differential privacy: $\frac{t^2}{2\sigma^2}$ is the CDP parameter of the Gaussian distribution.

**Lemma 15.4.** *Let $Z \leftarrow \mathrm{LLN}(\sigma)$ for $\sigma>0$. Let $s\in\mathbb{R}$ and $\alpha\in(1,\infty)$. Then*

$$\mathrm{D}_\alpha(Z\|Z+s) \leq \min\left\{\frac{1}{2}e^{3\sigma^2}s^2\alpha, e^{\frac{3}{2}\sigma^2}s\right\}.$$

The idea of proof this lemma is calculating the log density ratio and show that the ratio is bounded by $e^{\frac{3}{2}\sigma^2}s$ with probability 1. And when $\alpha$ is small, any pure DP mechanism satisfies the upper bound: $\frac{1}{2}e^{3\sigma^2}s^2$.

By combining lemma15.3 and lemma 15.4, we define $a$ as $\frac{t^2}{2\sigma^2}$ and $b$ as $\frac{1}{2}e^{3\sigma^2}s^2$, we can get the bound in proposition 15.1.

### 15.1.3 Examples: Releasing reciprocal and Private Argmax

Let f(D) be a counting query, define g(D) = 1/f(D). The global sensitivity of g(D) is $+\infty$ and the local sensitivity of g(D) is $\left|\frac{1}{f(D)} - \frac{1}{f(D)-1}\right| \leqslant O\left(\frac{1}{|f(D)|}\right)$. The smooth sensitivity of g(D) is $e^{-n\beta} \cdot \infty = \infty$. This is an example where the smooth sensitivity does not exist at all.

Suppose that we have a linear regression model and we can estimate $\hat{\theta}$ from the data $\hat{\theta} = \left(x^\top x\right)^{-1} x^\top \tilde{y}$. How close is $\tilde{x}^\top \hat{\theta}$ to $\tilde{x}^\top \theta^*$? This kind of standard deviation is proportional to $\sqrt{\tilde{x}\left(x^\top x\right)^{-1}\tilde{x}}$ .

Considering voting, model selection and top-k movie questions, for these tasks, we cannot use smooth sensitivity.

### 15.1.4 Release Stable Values without adding noise

Let's define "Dist2Instability" function:

$$d(x) = min\{d(x, x'') - 1\}$$

such that $x''$ satisfies that $f\left(x''\right) \neq f(x)$, which means that how many steps, or data points we need to add/remove before $x$ can be made into $x''$ such that the output of $x''$ is different from $f(x)$, where $f$ is the query we want to release. The global sensitivity of d(x) is 1. The mechanism will do the following:

1. $\hat{d} = d(x) + \text{Lap}\left(\frac{1}{\varepsilon}\right)$
2. if $\hat{d} > \frac{\log\frac{1}{\delta}}{\varepsilon}$, then output f(x);

   else if $\hat{d} \leq \frac{\log\frac{1}{\delta}}{\varepsilon}$, then output "$\perp$". We use the symbol "$\perp$" to denote nothing.

Then, we will show the privacy analysis of "Dist2Instability" by discussing two cases:

- Case A: $d(x) = 0 \Rightarrow \forall x'$ that is a neighbor of $x$, such that $f(x)' \neq f(x)$, then $d\left(x'\right) = 0$:

  $d(x) = 0 \Rightarrow \exists x''$ such that $d\left(x, x''\right) \leqslant 1$ and $f(x)'' \neq f(x)$. If $d(x, x') \leq 1$, then $d(x') = min\{d(x', x'') - 1\} = 0$.

  On data set x and $x'$, we are both adding Laplace noise to 0:

  $x : 0 + \text{Lap}\left(\frac{1}{\varepsilon}\right)$

  $x' : 0 + \text{Lap}\left(\frac{1}{\varepsilon}\right)$

  These two outputs should be identical unless $\text{Lap}\left(\frac{1}{\varepsilon}\right) > \frac{\text{Lap}\left(\frac{1}{\varepsilon}\right)}{\epsilon}$, which happens with probability less than $\delta$. So in the Case A, "Dist2Instability" satisfies $(0, \delta)$ - DP.

- Case B: $d(x) \neq 0$, then $f(x) = f(x')$, $\forall x'$ such that $d(x', x) = 1$. We need to discuss two cases:

  1. output is $\perp$. When $\hat{d} > \frac{\text{Lap}\left(\frac{1}{\varepsilon}\right)}{\epsilon}$, it is a post-processing of $\hat{d} = d(x) + \text{Lap}\left(\frac{1}{\varepsilon}\right)$
  2. output is $f(x)$. When $\hat{d} < \frac{\text{Lap}\left(\frac{1}{\varepsilon}\right)}{\epsilon}$, it is also a post-processing of $\hat{d} = d(x) + \text{Lap}\left(\frac{1}{\varepsilon}\right)$

  Then in Case B, "Dist2Instability" satisfies $\epsilon$ - DP.

Considering case A and B together, "Dist2Instability" satisfies $(\epsilon, \delta)$ - DP.

## 15.2 Propose-Test-Release

Firstly, we propose a bound $\beta$ on local-sensitivity and test the validity of this bound:

$$\hat{d} = d\left(x, \{x' : \mathrm{LS}_q\left(x'\right) > \beta\}\right) + \mathrm{Lap}(1/\varepsilon).$$

Then, we release the result: if $\hat{d} > \frac{\log \frac{1}{\delta}}{\varepsilon}$, return $q(x) + \mathrm{Lap}\left(\frac{\beta}{\varepsilon}\right)$; else return "$\perp$".

**Proposition 15.5.** *(propose-test-release [33]). For every query $q : X^n \to \mathbb{R}$ and $\varepsilon, \delta, \beta \geq 0$, the above algorithm is $(2\varepsilon, \delta)$-differentially private.*

### 15.2.1 The privacy analysis of PTR

We will show the privacy analysis of PTR by discussing two cases:

- Case 1: When the local sensitivity of dataset x is greater than $\beta$:

$$LS_q(x) > \beta \Rightarrow d\left(x, \{x'' : LS_q\left(x'\right) > \beta)\right) = 0$$
$$\hat{d}(x) = 0 + \mathrm{Lap}\left(\frac{1}{\varepsilon}\right) \quad \mathbb{P}\left(\hat{d}(x) > \frac{\mathrm{Lap}\,\frac{1}{\delta}}{\varepsilon}\right) \leq \delta$$

  Let's define $\hat{d}(x) > \frac{\mathrm{Lap}\,\frac{1}{\delta}}{\varepsilon}$ as event $E$. And the output space $S$ is $\mathbb{R} \cup \{\perp\}$, we have:

$$\begin{aligned}\mathbb{P}(M(x) \in S) &= \mathbb{P}\left(M(x) \in S \cap E^C\right) + \mathbb{P}(M(x) \in S \cap E) \\ &\leqslant \mathbb{P}(M(x) = " \perp ") + \delta \\ &\leqslant e^{\epsilon}\mathbb{P}\left(M(x) \in S \cap E^C\right) + \delta \\ &\leqslant e^{\epsilon}\mathbb{P}\left(M(x) \in S\right) + \delta\end{aligned}$$

  Then, it satisfies $(\epsilon, \delta)$-DP.

- Case 2: When the local sensitivity of dataset x is less or equal than $\beta$:

$$\begin{aligned}&LS(x) \leqslant \beta \\ &|q(x) - q\left(x'\right)| \leqslant \beta\end{aligned}$$

  The Laplace mechanism outputs $M(x) \in \mathbb{R} \cup \{\perp\}$. This is the adaptive composition of $\hat{d}$, which is a post-processing of q-DP mechanism; and $q(x) + \mathrm{Lap}\left(\frac{1}{\varepsilon}\right)$ and $q(x') + \mathrm{Lap}\left(\frac{1}{\varepsilon}\right)$ are $\epsilon$-indistinguishable. Then, it satisfies $(2\epsilon, 0)$-DP.

Considering case 1 and case 2 together, PTR satisfies $(2\epsilon, \delta)$-DP.

### 15.2.2 Releasing local sensitivity privately

How do we know what bound to propose? Let's say we want to estimate the number of triangles in a graph under Edge Differential Privacy. The global sensitivity is $n - 2$. The local sensitivity is bounded by the max degree of G. The local sensitivity has itself a global sensitivity that is 1: $\Delta_{LS_q} = 1$.

If we can provide a high probability upper bound of the local sensitivity, then we can add noise and it satisfies $(2\varepsilon, \delta)$-DP.

**Lemma 15.6.** *Let $\tilde{\Delta}_f(D)$ satisfies $\varepsilon$-DP and*

$$\mathbb{P}\left[\Delta_f(D) \geq \tilde{\Delta}_f(D)\right] \leq \delta$$

*Then $f(D) + \mathrm{Lap}\left(\tilde{\Delta}_f(D)/\epsilon\right)$ satisfies $(2\varepsilon, \delta)$-DP.*

*Proof.* Let's say $(y, \widetilde{\Delta})$ is the output of the algorithm, where $\widetilde{\Delta}$ is $\epsilon$-DP and $y = f(x) + \mathrm{Lap}\left(\frac{\widetilde{\Delta}}{\epsilon}\right)$. Let's define $E$ as $\left\{E \mid \widetilde{\Delta}_f > \Delta_f\right\}$:

$$\begin{aligned}
\mathbb{P}\left[(y, \widetilde{\Delta}) \in S_1 \times S_2 \mid x\right] &= \mathbb{P}_x\left[(y, \widetilde{\Delta}) \in S_1 \times S_2 \cap E\right] + \mathbb{P}_x\left[(y, \widetilde{\Delta}) \in S_1 \times S_2 \cap E^C\right] \\
&\leqslant \int_{\widetilde{\Delta} \in S_2 \cap E} \left[y \in S_1 \mid \widetilde{\Delta}\right] d\widetilde{\Delta} \cdot \mathbb{P}_x\left[\widetilde{\Delta} \in S_2 \cap E\right] + \delta \\
&\leqslant \int_{\widetilde{\Delta} \in S_2 \cap E} e^{\varepsilon} \mathbb{P}_{x'}\left[y \in S_1 \mid \tilde{\Delta}\right] d\tilde{\Delta} \cdot e^{\varepsilon} \mathbb{P}_{x'}\left[\widetilde{\Delta} \in S_2 \cap E\right] + \delta \\
&\leqslant e^{2\varepsilon} \mathbb{P}_{x'}\left[(y, \widetilde{\Delta}) \in S_1 \times (S_2 \cap E)\right] + \delta \\
&\leqslant e^{2\varepsilon} \mathbb{P}_{x'}\left[(y, \widetilde{\Delta}) \in S_1 \times S_2\right] + \delta
\end{aligned}$$

Then, it satisfies $(2\varepsilon, \delta)$-DP. □