

## DS102 - Discussion 12

Wednesday, 3rd December, 2019

In this section, we discuss differential privacy. First we recall its definition. For two databases  $S$  and  $S'$  which differ in only one entry (e.g. differing in one individual), an  $\epsilon$ -differentially private algorithm  $\mathcal{A}$  satisfies:

$$\mathbb{P}(\mathcal{A}(S) = a) \leq e^\epsilon \mathbb{P}(\mathcal{A}(S') = a),$$

for all points  $a$ . In words, the probability of seeing any given output of a differentially private algorithm doesn't change a lot by replacing only one entry in the input database.

We usually refer to databases that differ in only one entry as *neighboring* databases.

1. **Laplace mechanism.** One of the most widely used mechanisms for differential privacy is the *Laplace mechanism*. The idea is as follows. Suppose that we want to report a statistic  $f(\cdot)$ , which takes as input a database. For example,  $S$  could be a database with the salaries of all residents of Berkeley, and  $f(S)$  could be the average salary in  $S$ . Denote by  $S$  and  $S'$  generic neighboring databases (meaning they differ in only one entry). Define the sensitivity of  $f$  as:

$$\Delta_f = \max_{\text{neighboring } S, S'} |f(S) - f(S')|.$$

The Laplace mechanism reports  $\mathcal{A}_{\text{Lap}}(S) = f(S) + \xi_\epsilon$ , where  $\xi_\epsilon$  is distributed according to the zero-mean Laplace distribution with parameter  $\frac{\Delta_f}{\epsilon}$ , denoted  $\text{Lap}(0, \frac{\Delta_f}{\epsilon})$ . The Laplace distribution  $\text{Lap}(\mu, b)$  is given by the following density:

$$p(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}.$$

The Laplace distribution is essentially a two-sided exponential distribution.

- (a) Prove that the Laplace mechanism is  $\epsilon$ -differentially private. More precisely, show that for all  $S'$  that are neighboring to our database  $S$ , we have

$$\frac{\mathbb{P}(\mathcal{A}_{\text{Lap}}(S) = a)}{\mathbb{P}(\mathcal{A}_{\text{Lap}}(S') = a)} \leq e^\epsilon.$$

**Solution:** Since  $\xi_\epsilon \sim \text{Lap}(0, \frac{\Delta_f}{\epsilon})$ , adding  $f(S)$  simply shifts the mean, so  $\mathcal{A}_{\text{Lap}}(S) = f(S) + \xi_\epsilon \sim \text{Lap}(f(S), \frac{\Delta_f}{\epsilon})$ . Similarly, for any neighboring set  $S'$ ,  $\mathcal{A}_{\text{Lap}}(S') = f(S') + \xi_\epsilon \sim \text{Lap}(f(S'), \frac{\Delta_f}{\epsilon})$ . We want to show that the ratio of

these two densities is bounded by  $e^\epsilon$ . At point  $x \in \mathbb{R}$ , the density ratio is:

$$\begin{aligned} \frac{\mathbb{P}(\mathcal{A}_{\text{Lap}}(S) = a)}{\mathbb{P}(\mathcal{A}_{\text{Lap}}(S') = a)} &= \frac{\epsilon/2\Delta_f e^{-\frac{\epsilon|x-f(S)|}{\Delta_f}}}{\epsilon/2\Delta_f e^{-\frac{\epsilon|x-f(S')|}{\Delta_f}}} \\ &= \frac{e^{-\frac{\epsilon|x-f(S)|}{\Delta_f}}}{e^{-\frac{\epsilon|x-f(S')|}{\Delta_f}}} \\ &= e^{\frac{-\epsilon(|x-f(S)|+|x-f(S')|)}{\Delta_f}}. \end{aligned}$$

By triangle inequality, we have  $|x - f(S')| - |x - f(S)| \leq |f(S) - f(S')|$ , and moreover this is upper bounded by  $\Delta_f$ , by definition of sensitivity. Therefore:

$$e^{\frac{-\epsilon(|x-f(S)|+|x-f(S')|)}{\Delta_f}} \leq e^{\frac{\epsilon\Delta_f}{\Delta_f}} = e^\epsilon.$$

- (b) In part (a) we convinced ourselves that the Laplace mechanism indeed ensures privacy. However, privacy alone is easy to ensure - one can always report random noise. To also have utility from the reported values, we have to consider a trade-off between privacy and *accuracy*. Accuracy means that  $\mathcal{A}_{\text{Lap}}(S)$  is actually close to  $f(S)$  with high probability.

Using the fact that  $X \sim \text{Lap}(0, b)$  satisfies:

$$\mathbb{P}(|X| \geq t) \leq 2e^{-\frac{t}{b}},$$

prove that the Laplace mechanism also enjoys nice accuracy guarantees:

$$\mathbb{P}(|\mathcal{A}_{\text{Lap}}(S) - f(S)| \geq t) \leq 2e^{-\frac{t\epsilon}{\Delta_f}}.$$

**Solution:** Since  $\mathcal{A}_{\text{Lap}}(S) - f(S) = f(S) + \xi_\epsilon - f(S) = \xi_\epsilon \sim \text{Lap}(0, \frac{\Delta_f}{\epsilon})$ , we can apply the above inequality with  $b = \frac{\Delta_f}{\epsilon}$  to get:

$$\mathbb{P}(|\mathcal{A}_{\text{Lap}}(S) - f(S)| \geq t) \leq 2e^{-\frac{t\epsilon}{\Delta_f}}.$$

- (c) What can you conclude about the relationship between sensitivity  $\Delta_f$  and accuracy, for a fixed level of privacy  $\epsilon$ ? Does this make intuitive sense?

**Solution:** If  $\Delta_f$  is large, we have to add large amounts of noise to ensure privacy. For this reason, the reported values will also be a lot less accurate for

large  $\Delta_f$  (because  $e^{-\frac{t\epsilon}{\Delta_f}}$  is increasing in  $\Delta_f$ ). This makes sense, because if we have very noticeable outliers in our data set - e.g. we want to report the average salary and we have the richest person in the world in our database - to make the result insensitive to replacing the richest person in the world someone whose income is 0, our report has to be very noisy and hence inaccurate.

- (d) Suppose you want to report the average salary, i.e.  $f(S) = \frac{1}{n} \sum_{i=1}^n s_i$ , where  $s_i$  is the salary of the  $i$ -th individual in the database. Moreover, suppose that all salaries are in the range  $[0, M]$ . What is an appropriate parameter of the Laplace mechanism, if we want to report the average salary in an  $\epsilon$ -differentially private way? What is the accuracy guarantee of this mechanism?

**Solution:** The sensitivity of  $f$  is  $\Delta_f = \frac{M}{n}$  (because we can replace someone with salary  $M$  with someone with salary 0, or vice versa). Therefore, we need noise  $\xi_\epsilon \sim \text{Lap}(\frac{M}{n\epsilon})$ . The accuracy is worse than  $t$  with probability at most  $2e^{-\frac{t\epsilon n}{M}}$ .

2. **Post-processing of differential privacy.** An important property of differential privacy is that it is preserved under post processing: if  $\mathcal{A}(S)$  is an  $\epsilon$ -differentially private reported statistic, then  $g(\mathcal{A}(S))$  is still differentially private, for any function  $g$ . Prove this fact.

**Solution:** Define  $T_c := \{a : g(a) = c\}$ . Because  $\mathcal{A}$  is  $\epsilon$ -differentially private, we know

$$\mathbb{P}(\mathcal{A}(S) = a) \leq e^\epsilon \mathbb{P}(\mathcal{A}(S') = a),$$

We use this to get

$$\begin{aligned} \mathbb{P}(g(\mathcal{A}(S)) = c) &= \mathbb{P}(\mathcal{A}(S) \in T_c) \\ &\leq e^\epsilon \mathbb{P}(\mathcal{A}(S') \in T_c) \\ &= e^\epsilon \mathbb{P}(g(\mathcal{A}(S')) = c). \end{aligned}$$