The Algorithmic Foundations of Adaptive Data Analysis November 15, 2020

Lecture 11: Privacy and Adaptive Analysis

Lecturer: Adam Smith *Scribe: Adam Smith*

1 Differential Privacy

Distributional stability notions can be interpreted as providing various notions of privacy.

Definition 1 (Differential Privacy) *A randomized algorithm $M : \mathcal{X}^n \to \mathcal{Y}$ is* (ϵ, δ)-differentially private *if for all pairs of data sets $\mathbf{s}, \mathbf{s}' \in \mathcal{X}^n$ that differ in one input, for all events $E \subseteq \mathcal{Y}$,*

$$\Pr(M(\mathbf{s}) \in E) \leq e^\epsilon \Pr(M(\mathbf{s}') \in E) + \delta \,.$$

Recall that stability in $d_\diamond$ is the same as $(\epsilon, 0)$-differential privacy. The more general definition, with $\delta > 0$, is based on a two-parameter measure of closeness of distributions: we say probability measures P and Q are (ϵ, δ)*-indistinguishable* if

$$P(E) \leq e^\epsilon Q(E) + \delta \quad \text{and} \quad Q(E) \leq e^\epsilon P(E) + \delta \quad \text{for all events } E\,.$$

Informally, differential privacy ensures that an attacker viewing the output of M will draw approximately the same conclusion no matter what the data for person i are (assuming that the entries of the data set correspond one-to-one with a specific set of people).

1.1 Notions of adjacency

So far in this class, we've looked at how robust algorithms are to the replacement of a single data point. This notion of change corresponds to a natural metric on data set: the *Hamming distance* between two lists $\mathbf{s}, \mathbf{s}' \in \mathcal{X}^n$ is the number of entries of $\mathbf{s}$ that must be changed to obtain $\mathbf{s}'$. Two data sets are adjacent (or "neighboring") if their Hammin distance is 1.

Differential privacy is usually defined with respect to a different metric, *set difference*. Specifically, we view data sets as finite multisets of $\mathbf{s}$ (where points can appear with mutiplicities other than 1) or, equivalently as "characteristic functions" $\mathbf{s} : \mathcal{X} \to \mathbb{N}$. The set difference metric between $\mathbf{s}$ and $\mathbf{s}'$ is the size of the symmetric difference $|\mathbf{s} \triangle \mathbf{s}'|$ (or the ℓ_1 distance between their characteristic functions). Requiring the (ϵ, δ)-indistinguishability of distributions on outputs of inputs adjacent in this metric corresponds to requiring that an adversary viewing the output *not be able to distinguish whether or not a particular person's data were used in the computation.*

Depending on the context, other notions of adjacency may make sense for privacy (for example, when the data is a social network, it might make sense to consider what happens when an arbitrary vertex is inserted or deleted from the network). For this course, we will largely stick to the Hamming distance-based version (Definition 1).

Why does this definition correspond to "privacy" in the context of statistical data? Imagine that the event E captures those outputs that lead to a "bad" outcome for a user i (being denied health insurance, for example). Then differential privacy ensures that the bad outcomes occur with at most $e^\epsilon \approx 1 + \epsilon$ times higher probability when person i's real data are used.

In particular, differentially private algorithms resist "membership inference attacks" (also called tracing attacks) that seek to determine if a particular person is in some data set or not. (Note that membership alone can be sensitive! Think, for example, of the data from a clinical study of a particular disease, or a study of drug addicts, or a study of undocumented immigrants.

For the purposes of this course, the relevant fact is that differentially private algorithms are distributionally stable, and hence provide generalization guarantees.

1.2 Basic Mechanisms

1.2.1 Noise Addition

Recall the Laplace mechanism from Lecture 7, which we formulate here slightly more generally:

Definition 2 *A query $q : \mathcal{X}^n \to \mathbb{R}^d$ has sensitivity Δ in the ℓ_1 norm if for all $x_1, \ldots, x_n \in \mathcal{X}$, all indices i, and all $x_i' \in \mathcal{X}$:*

$$\|q(x_1, \ldots, x_n) - q(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)\|_1 \leq c$$

Note that statistical queries are Δ-sensitive for $\Delta = 1/n$.

Algorithm 1: Laplace mechanism$(\epsilon, \mathbf{s})$

```
**Input:** Data set $\mathbf{s} = (x_1, ..., x_n) \in \mathcal{X}^n$ and parameter $\epsilon > 0$.
1 Receive a $\Delta$-sensitive query $q : \mathcal{X} \to [0, 1]$ from analyst;
2 **return** $\frac{1}{n} \sum_{i=1}^{n} q(x_i) + Z_1, ..., Z_d$ where $Z_i \sim \text{Lap}(0, \frac{\Delta}{\epsilon})$ are i.i.d.
```

Lemma 3 *The Laplace mechanism is ϵ-differentially private. If repeated k times, the Laplace mechanism is $k\epsilon$-dfferentially private.*

Exercise 1 *Show that the Gaussian mechanism is (ϵ, δ)-differentially private when $\sigma \geq \frac{2c\sqrt{\ln(1/\delta)}}{\epsilon}$, where c is the sensitivity of the query in the ℓ_2 norm (as opposed to the ℓ_1 norm).*

1.2.2 Exponential Sampling

What other interesting mechanisms satisfy the definition? The Laplace and Gaussian mechanism are useful for answering numerical queries that have low sensitivity, but what about when adding noise makes no sense? The *exponential mechanism* is the natural starting point for designing differentialy private algorithms.

We'll motivate the mechanism with two selection problems:

Example 1 *Suppose we gather data from a clinical study that considered a set of d medical conditions, which we will abstractly number $\{1, ..., d\}$. Our data set consists of, for each person i in a clinical study, the subset $x_i \subseteq \{1, ..., d\}$ of medical conditions experience by individual i.*

We wish to know the (approximately) most common medical condition. For each condition j, let $q(j; \mathbf{s}) = |\{i : j \in x_i\}|$. We want to learn an index $\hat{\jmath}$ such that $q(\hat{\jmath}; \mathbf{s})$ is as large as possible.

Example 2 (Prices of a digital good) *Adele has recorded her new hit, "Goodbye", and wants to sell it online. In a market study, she collects from n people the price $x_i \in [0, 1]$ they would be willing to pay for a download of the song. Assuming that respondents answered truthfully, then a reasonable estimate for the revenue Adele would get from selling the download at price p is*

$$q(p; \mathbf{s}) = p \cdot \#\{i : x_i \geq p\}\,.$$

Adele would like to learn a price $\hat{p} \in \{\$0.01, \$0.02, \ldots \$0.99\}$ such that $q(\hat{p}; \mathbf{s})$ is as large as possible.

These examples have a lot in common. We can fit both these problems into a general framework, specified by

- A set $\mathcal{Y}$ of possible outputs (conditions and prices in the examples above).
- A score function $q : \mathcal{Y} \times \mathcal{X}^n \to \mathbb{R}$ which measures the "goodness" of each output for a data set. Given $\mathbf{s} \in \mathcal{X}^n$, our goal is to find $y \in \mathcal{Y}$ which approximately maximizes $q(y; \mathbf{s})$. (When $\mathcal{Y}$ is finite, we can also think of q as a collection of $\mathcal{Y}$ separate low-sensitivity queries.)
- A sensitivity bound $\Delta > 0$ such that $q(y; \cdot)$ is Δ-sensitive for every y. That is,

$$\sup_{y \in \mathcal{Y}} \sup_{\substack{\mathbf{s}, \mathbf{s}' \in \mathcal{X}^n \\ \text{adjacent}}} |q(y; \mathbf{s}) - q(y; \mathbf{s}')| \leq \Delta\,. \tag{1}$$

We get the following algorithm:

```
Algorithm 2: Exponential Mechanism $M_{EM}(\mathbf{s}, score, \Delta, \epsilon)$
  Input: Assume that $q(y;\cdot)$ is $\Delta$-sensitive for every $y \in \mathcal{Y}$.
  Select $Y$ such that $\Pr(Y = y) \propto \exp\left(\frac{\epsilon}{2\Delta} q(y;\mathbf{s})\right)$;
  return $Y$;
```

Before we analyze privacy, it is worth asking under what conditions this mechanism is well defined. When $\mathcal{Y}$ is finite the algorithm is well-defined since we can set

$$P(Y = y) = \frac{e^{\frac{\epsilon}{2\Delta} q(y;\mathbf{s})}}{\sum_{y' \in \mathcal{Y}} e^{\frac{\epsilon}{2\Delta} q(y';\mathbf{s})}} .$$

In fact, the mechanism makes sense over any domain as long as there is a base measure such that $\int_{y \in \mathcal{Y}} \exp\left(\frac{\epsilon}{2\Delta} q(y;\mathbf{s})\right) dy$ is well-defined and finite for every possible data set $\mathbf{s}$. We will see an example further below.

Theorem 4 *If score is Δ-sensitive (i.e., satisfies* (1)*) then the exponential mechanism is ϵ-differentially private.*

Proof Assume for simplicity that $\mathcal{Y}$ is finite. For any output y and data set $\mathbf{s}$ we have $P(y|\mathbf{s}) = \frac{e^{\frac{\epsilon}{2\Delta} q(y;\mathbf{s})}}{\sum_{y' \in \mathcal{Y}} e^{\frac{\epsilon}{2\Delta} q(y';\mathbf{s})}}$. Let $\mathbf{s}'$ be a data set adjacent to $\mathbf{s}$. Since the sensitivity of $q(y;\cdot)$ is at most Δ, we have

$$\frac{e^{\frac{\epsilon}{2\Delta} q(y;\mathbf{s})}}{e^{\frac{\epsilon}{2\Delta} q(y;\mathbf{s}')}} = \exp\left(\frac{\epsilon}{2\Delta}(q(y;\mathbf{s}) - q(y;\mathbf{s}'))\right) \leq e^{\epsilon/2}$$

and similarly, for the normalizing constants,

$$\frac{\sum_{y' \in \mathcal{Y}} e^{\frac{\epsilon}{2\Delta} q(y';\mathbf{s}')}}{\sum_{y' \in \mathcal{Y}} e^{\frac{\epsilon}{2\Delta} q(y';\mathbf{s})}} \leq \sup_{y'} \left(\exp\left(\frac{\epsilon}{2\Delta}(q(y';\mathbf{s}') - q(y';\mathbf{s}))\right)\right) \leq e^{\epsilon/2} .$$

Thus the ratio $\frac{Pr(y|\mathbf{s})}{P(y|\mathbf{s}')}$ is at most $e^{\epsilon/2} \cdot e^{\epsilon/2} = e^{\epsilon}$. The case of an infinite domain is similar, with integrals over to the base measure replacing sums. ■

When the domain is finite, it is often more convenient to work with a another algorithm which behaves very similarly:

```
Algorithm 3: Report-Noisy-Max $M_{RNM}(\mathbf{s}, score, \Delta, \epsilon)$
  Input: Assume that $q(y;\cdot)$ is $\Delta$-sensitive for every $y \in \mathcal{Y}$, and $\mathcal{Y} = \{1, ..., d\}$ is finite
  Select $Z_1, ..., Z_d \sim \text{Exp}(2\Delta/\epsilon)$ i.i.d. ;
  return $\arg\max_{y \in \{1,...,d\}} \left(q(y;\mathbf{s}) + Z_y\right)$;
```

This algorithm is generally much easier to implement than the exponential mechanism, since it does not require explicitly computing any probabilities and can make use of standard libraries for sampling from the exponential distribution. It satisfies a very similar guarantee to the exponential mechanism.

Exercise 2 *Show that report noisy max is ϵ-differentially private. [*Hint: *Consider two outputs a, b. For a fixed input $\mathbf{s}$, what is $\frac{P(a|\mathbf{s})}{P(b|\mathbf{s})}$?]*

In addition to making implementation easier, the utility analysis of report-noisy-max is more intuitive. For one thing, the amount of noise added to each score is just Δ/ϵ, independent of the number of possible outputs d. [1]

Lemma 5 (Tail Bounds for Exponential Distributions)

1. *If $Z \sim Exp(\lambda)$, then $\Pr(Z \geq t\lambda) = e^{-t}$ for all $t \geq 0$.*
2. *If $Z_1, ..., Z_d \sim Exp(\lambda)$ i.i.d,, then $\Pr(\max_{i=1}^{d} Z_i > \lambda(\ln(d) + t)) = e^{-t}$ for all $t \geq 0$.*

[1] In contrast, if we were to release *all* of the noisy scores, then the Laplace mechanism would tell us to add noise $d \cdot \Delta/\epsilon$ to each entry since the ℓ_1 sensitivity of the vector of answers is $d \cdot \Delta$.

Note that if Y_i are independent Laplace random variables with $Y_i \sim \text{Lap}(\mu_i, \lambda_i)$ and $Z_i = |Y_i - \mu_i|$, then the Z_i's will be exponentially distributed with parameter λ and so Lemma 5 above applies.
Proof The first part follows from a direct computation of the CDF:

$$Pr(Z > \lambda t) = \int_{y \geq \lambda t} \tfrac{1}{\lambda} e^{-y/\lambda} dy = \tfrac{1}{\lambda} \left[-\lambda e^{-y/\lambda} \right]_{y=\lambda t}^{\infty} = e^{-t} \,.$$

The second part follows by a union bound: the probability that any particular Z_i exceeds $\lambda(\ln(d) + t)$ is $\frac{e^{-t}}{d}$ by part 1, so the probability that any of the Z_i's exceeds the bound is at most e^{-t}. ■

We can also use Lemma 5 to prove the following:

Theorem 6 *If $q(y;\cdot)$ is Δ-sensitive for every $y \in \{1, ..., d\}$, then for every data set $\mathbf{s}$ in $\mathcal{X}^n$ and every $t > 0$, the output of report-noisy-max $Y \leftarrow M_{RNM}(\mathbf{s}, score, \Delta, \epsilon)$ satisfies*

$$\Pr\left(q_{max}(\mathbf{s}) - q(Y, \mathbf{s}) \geq \frac{4\Delta(\ln(d) + t)}{\epsilon} \right) \leq e^{-t} \,, \ \textit{where } q_{max}(\mathbf{s}) = \max_{y=1}^{d} q(y; \mathbf{s}) \,,$$

and

$$\mathbb{E}\left(q_{max}(\mathbf{s}) - q(Y, \mathbf{s}) \right) = O(\Delta \ln(d)/\epsilon) \,.$$

One consequence of this is that if we have a collection of d statistical queries, we can get an approximate maximizer from the collection with error that scales as $\ln(d)/(n\epsilon)$. In the next lecture, we will see an "online" version of this result.

Exercise 3 *(1) Prove Theorem 6. (2) Show that the exponential mechanism, M_{EM} satisfies a similar guarantee to Theorem 6 (possibly with a larger constant than 4).*

2 Notes

The Laplace and exponential mechanisms (followed closely by the Gaussian mechanism and the Sparse Vector algorithm from next lecture) are the two most common building blocks of differentially private algorithms. The Laplace mechanism is from [DMNS06], while the exponential mechanism is due to McSherry and Talwar [MT07].

The privacy of the report-noisy-max mechanism appears in several places, for example in Chen et al. [CCK+11]. Report noisy max can be instantiated with several other noise distributions (for example, Gumbel or Laplace noise), and will satisfy similar guarantees, even though the exact distribution on outputs changes. It turns out that instantiating report noisy max with the Gumbel distribution leads to an algorithm that samples *exactly* from the exponential mechanism distribution. This fact is folklore in machine learning, and known as the "Gumbel Max Trick".

References

[CCK+11] Yiling Chen, Stephen Chong, Ian A. Kash, Tal Moran, and Salil P. Vadhan. Truthful mechanisms for agents that value privacy. *CoRR*, abs/1111.5472, 2011.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, pages 265–284. Springer, 2006.

[MT07] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS*, 2007.