CS839: Differential Privacy and Learning January 22, 2026

Lecture 2: Privacy Basics I

Instructor: Gavin Brown *Scribe:* Akhil Vanukuri

In this lecture, we will develop some tools that will be helpful in formalizing the notion of what it means for an algorithm to be differentially private while being accurate. We will also look at two concrete algorithms and argue about the differential privacy guarantees that they provide.

1 Preliminaries

- We will assume that all the events and distributions are defined over appropriate supports such that they are well-defined. We will ignore measure theory whenever possible.
- The words mechanism and algorithm will be used interchangeably.

1.1 Definitions

For any datasets $x, x' \in \mathcal{X}^n$, x is said to be *adjacent* to x' (denoted by $x \sim x'$) if

$$(\exists i^* : \forall i \neq i^*, x_i = x'_i)$$

where $i, i^* \in [n]$ and $\mathcal{X}$ is the domain over which the datasets are defined.

Definition 1.1 ((ϵ, δ)-Differentially Private)**.** An algorithm $\mathcal{A}$ is $(\epsilon, \delta) - \mathsf{DP}$ (Differentially Private) if $\forall\ x \sim x'$ and all events E,

$$\Pr[\mathcal{A}(x) \in E] \leq e^{\epsilon} \cdot \Pr[\mathcal{A}(x' \in E)] + \delta$$

When $\delta = 0$, we will use the short-hand ϵ-DP for $(\epsilon, 0)$-DP.

Definition 1.2 (Privacy Loss)**.** We denote the privacy loss for an algorithm $\mathcal{A}$, datasets x, x' and, output y as $L_{\mathcal{A}}^{x \to x'}(y)$ where $x \sim x'$ and,

$$L_{\mathcal{A}}^{x \to x'}(y) := \ln \left(\frac{\Pr[\mathcal{A}(x) = y]}{\Pr[\mathcal{A}(x') = y]} \right)$$

Note: $\mathcal{A}$ is ϵ-DP $\iff$ $\forall x \sim x'$, $\forall y$, $L_{\mathcal{A}}^{x \to x'}(y) \leq \epsilon$.

Definition 1.3 (Laplace distribution)**.** The probabilistic density function (denoted by p) of a Laplace distribution characterized by the mean μ and the scale b is defined as

$$p(z) := \frac{1}{2b} \exp \frac{-|z - \mu|}{b}$$

We will use $\mathsf{Lap}(\mu, b)$ to refer to this distribution and, the shorthand $\mathsf{Lap}(\mathsf{b})$ when $\mu = 0$.

2 Properties of DP algorithms

We will state some general properties of DP algorithms.

Claim 2.1 (Immunity to Post-processing)**.** *If an algorithm $\mathcal{A} : \mathcal{X}^n \to \mathcal{Y}$ is (ϵ, δ)-DP and $f : \mathcal{Y} \to \mathcal{Z}$ is a randomized function then $f \circ \mathcal{A}$[1] is a $(\epsilon, \delta) - \mathsf{DP}$.*

Proof. Let the function f be defined as $f : \mathcal{Y} \times \mathcal{R} \to \mathcal{Z}$ to factor in it's randomness. Let $\mathcal{R}$ be the domain from which the random coins are sampled. Fix any two datasets x, x' such that $x \sim x'$.

Observe that for any event $E \subseteq \mathcal{Z}$,

$$\begin{aligned}
\Pr[f(\mathcal{A}(x)) \in E] &= \sum_{r \in \mathcal{R}} \Pr[R = r] \cdot \Pr\left[\frac{f(\mathcal{A}(x), R = r) \in E}{R = r}\right] \\
&= \sum_{r \in \mathcal{R}} \Pr[R = r] \cdot \Pr\Big[f(\mathcal{A}(x), r) \in E\Big] \\
&= \mathbb{E}\Big[\Pr\Big[f(\mathcal{A}(x), r) \in E\Big]\Big] \\
&\leq \mathbb{E}\Big[e^{\epsilon} \cdot \Pr\Big[f(\mathcal{A}(x'), r) \in E\Big] + \delta\Big] \ (\because \ \mathcal{A} \text{ satisfies } (\epsilon, \delta) - \mathsf{DP}) \\
&\leq e^{\epsilon} \cdot \mathbb{E}\Big[\Pr\Big[f(\mathcal{A}(x'), r) \in E\Big]\Big] + \delta \\
&\leq e^{\epsilon} \cdot \Pr\Big[f(\mathcal{A}(x')) \in E\Big] + \delta
\end{aligned}$$

Note: The probability above is taken over the random coins of $\mathcal{A}$ and f and R is a random variable that takes the value of random coins of f. □

Claim 2.2 (Group Privacy)**.** *Suppose $\mathcal{A}$ is $(\epsilon, \delta) - \mathsf{DP}$ and let x, x' differ in at most k entries then $\mathcal{A}(x) \approx_{(\epsilon', \delta')} \mathcal{A}(x')$ for $\epsilon' = k \cdot \epsilon$ and $\delta' = k \cdot e^{\epsilon \cdot k} \cdot \delta$.*

Proof. Proof left as an exercise. □

Claim 2.3 (Basic Composition)**.** *Let $\mathcal{A}_1, \ldots, \mathcal{A}_k$ be independent $(\epsilon, \delta) - \mathsf{DP}$ algorithms defined over the same domain of datasets, then $\mathcal{A}(x) := (\mathcal{A}_1(x), \ldots, \mathcal{A}_k(x))$ is $(k\epsilon, k\delta) - \mathsf{DP}$.*

Proof. When $\delta = 0$ and the algorithms are independent, this result directly follows from writing down the definition of privacy loss. The complete proof is left as an exercise. □

3 Basic DP algorithms

In this section, we will define and analyze some basic DP algorithms.

[1] $f \circ \mathcal{A}(x) := f(A(x))$

3.1 Randomized Response

This is the simplest DP algorithm and it operates on a single bit ("w.p." stands for *with probability*, usually taken over the random coins of the algorithm).

Algorithm 1 Randomized Response ($\mathsf{RP}_\epsilon(x)$)

Input: $x \in \{0,1\}$, $\epsilon > 0$
Returns: $\tilde{x} \in \{0,1\}$

1: $\tilde{x} = \begin{cases} x & \text{w.p } \frac{e^\epsilon}{1+e^\epsilon} \\ 1-x & \text{w.p } \frac{1}{1+e^\epsilon} \end{cases}$

Claim 3.1. $\mathsf{RP}_\epsilon(x)$ *is* ϵ-DP

Proof. Fix any $x, x' \in \mathcal{X}^n$ such that $x \sim x'$. And, because $x, x' \in \{0,1\}$, w.l.o.g. we will assume $x = 0$ and $x' = 1$.

$$\begin{aligned} \Pr[\mathsf{RP}_\epsilon(0) = 0] &= \frac{e^\epsilon}{1+e^\epsilon} \\ &= \frac{1}{1+e^\epsilon} \cdot e^\epsilon \\ &= \Pr[\mathsf{RP}_\epsilon(1) = 0] \cdot e^\epsilon \end{aligned}$$

Similarly,

$$\begin{aligned} \Pr[\mathsf{RP}_\epsilon(0) = 1] &= \frac{1}{1+e^\epsilon} \\ &\leq \frac{1}{1+e^\epsilon} \cdot e^\epsilon \\ &\leq \Pr[\mathsf{RP}_\epsilon(1) = 1] \cdot e^\epsilon \end{aligned}$$

Therefore, $\mathsf{RP}_\epsilon(x)$ is ϵ-DP. □

3.2 Laplace Mechanism

Today, we will introduce the Laplace mechanism for a specific statistical task (mean estimation of Bernoulli trials). In a later class, we will discuss how it can applied quite generally.

For $n \in \mathbb{N}$, given $x_1, \ldots, x_n \xleftarrow{\text{i.i.d}} \mathsf{Ber}(p)$ where $p \in [0,1]$ [2]. The task of the Laplace mechanism is to predict the value of p.

Algorithm 2 Laplace Mechanism

Input: $x_1, \ldots, x_n \xleftarrow{\text{i.i.d}} \mathsf{Ber}(p)$, $\epsilon > 0$
Returns: $\tilde{p} \in [0,1]$

1: $\hat{s} := \sum_{i \in [n]} x_i$
2: $\tilde{s} := \hat{s} + z$ where $z \leftarrow \mathsf{Lap}(\frac{1}{\epsilon})$
3: $\tilde{p} := \frac{1}{n} \cdot \tilde{s}$

[2] Ber stands for the Bernoulli distribution with parameter p

3.2.1 Privacy Analysis

Claim 3.2. *Laplace mechanism is* $\epsilon - \mathsf{DP}$

Proof. Fix any $x, x' \in \mathcal{X}^n$ such that $x \sim x'$ where $x := (x_1, \ldots, x_n)$ and $x' := (x'_1, \ldots x'_n)$.

Let $E \subseteq \mathbb{R}$ be any event and $y \in E$. W.l.o.g., assume that $x_1 \neq x'_1$.

By claim 2.1, it suffices to analyze $\tilde{s}$ as the output of the above algorithm as $\tilde{p}$ depends only on the value of $\tilde{s}$.

$$\begin{aligned}
\Pr_{\tilde{s} \leftarrow \mathcal{A}(x)}[\tilde{s} = y] &= \Pr[\hat{s} + z = y] \\
&= \Pr[z = y - \hat{s}] \\
&= \frac{\epsilon}{2} \cdot \exp\big(- \epsilon \cdot |y - \sum_{i \in [n]} x_i| \big) \\
&= \frac{\epsilon}{2} \cdot \exp\big(- \epsilon \cdot |y - \sum_{i \in [n]} x_i + (x'_1 - x'_1)| \big) \\
&= \frac{\epsilon}{2} \cdot \exp\big(- \epsilon \cdot |\big(y - \sum_{i \in [n]} x'_i\big) + \big(x'_1 - x_1\big)| \big) \\
&\leq \frac{\epsilon}{2} \cdot \exp\big(- \epsilon |y - \sum_{i \in [n]} x'_i| + \epsilon |x'_1 - x_1| \big) \ (\because \forall a, b \in \mathbb{R},\ |a + b| > |a| - |b|) \\
&\leq \frac{\epsilon}{2} \cdot \exp\big(- \epsilon |y - \sum_{i \in [n]} x'_i| \big) \cdot \exp\big(\epsilon |x'_1 - x_1| \big) \\
&\leq \frac{\epsilon}{2} \cdot \exp\big(- \epsilon |y - \sum_{i \in [n]} x'_i| \big) \cdot \exp\big(\epsilon \big) (\because |x'_1 - x_1| = 1, \text{Bernoulli distribution variables}) \\
&\leq \Pr_{\tilde{s} \leftarrow \mathcal{A}(x')}[\tilde{s} = y] \cdot e^{\epsilon}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr_x[\tilde{s} \in E] &= \int_{y \in E} \Pr_x[\tilde{s} = y]\, dx \\
&\leq \int_{y \in E} \Pr_{x'}[\tilde{s} = y] \cdot e^{\epsilon}\, dx \ \text{(From the above)} \\
&\leq e^{\epsilon} \cdot \Pr_{x'}[\tilde{s} \in E]
\end{aligned}$$

□

3.2.2 Accuracy Analysis

Claim 3.3. *Laplace mechanism satisfies* $\mathbb{E}\big[(p - \tilde{p})^2\big] \leq \frac{p \cdot (1-p)}{n} + \frac{2}{\epsilon^2 \cdot n^2}$

Proof.

$$
\begin{aligned}
\mathbb{E}\big[(p-\tilde{p})^2\big] &= \mathbb{E}\big[\left(p-\frac{\tilde{s}}{n}\right)^2\big] \\
&= \mathbb{E}\left[\left(p-\frac{(\tilde{s}+\hat{s}-\hat{s})}{n}\right)^2\right] \\
&= \mathbb{E}\left[\left(p-\frac{(\hat{s}+(\tilde{s}-\hat{s}))}{n}\right)^2\right] \\
&= \mathbb{E}\left[\left(p-\frac{(\hat{s}+z)}{n}\right)^2\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[\left(np-\hat{s}-z\right)^2\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[\left(np-\hat{s}\right)^2\right] - \frac{1}{n^2}\mathbb{E}\left[2z(np-\hat{s})\right] + \frac{1}{n^2}\mathbb{E}[z^2] \\
&= \frac{1}{n^2}\mathbb{E}\left[\left(np-\hat{s}\right)^2\right] + \frac{1}{n^2}\mathbb{E}[z^2] \quad (\because \mathbb{E}[np-\hat{s}]=0) \\
&= \left(\mathsf{Variance\ of\ Bin}(n,p) + \mathsf{Variance\ of\ Lap}(\frac{1}{\epsilon})\right)\cdot\frac{1}{n^2} \\
&= \frac{np\cdot(1-p)}{n^2} + \frac{2}{\epsilon^2\cdot n^2} \\
&= \frac{p\cdot(1-p)}{n} + \frac{2}{\epsilon^2\cdot n^2}
\end{aligned}
$$

□

Note: The term $\frac{p\cdot(1-p)}{n}$ quantifies the sampling error[3] and the term $\frac{2}{\epsilon^2\cdot n^2}$ quantifies the cost of privacy[4]

We should have mixed feelings when interpreting this result. The good news: for any fixed distribution, the noise from privacy will approach zero much more quickly than the sampling error ($1/n^2$ versus $1/n$). However, the bad news is that when the distribution is heavily biased (i.e., $p \approx 0$ or $p \approx 1$), for any fixed n the privacy error could be much much larger than the sampling error. One of the goals of this class will be to develop estimators that better adapt to the specific solution.

[3]Because this term is due to the Bernoulli noise p

[4]Because this term is entirely due to the Laplacian error term z which is introduced for privacy.