# DIFFERENTIAL PRIVACY

Dinh Huu Nguyen, 06/25/2019

**Abstract**: a review of differential privacy and its application to machine learning.

Contents

## 0. Symbols and Terms

| | |
|---|---|
| $1, \dots, k$ | indexed by $f$ |
| $1, \dots, l$ | indexed by $g$ |
| $1, \dots, m$ | indexed by $h$ |
| $1, \dots, n$ | indexed by $i$ |
| $1, \dots, o$ | indexed by $j$ |
| $S$ | subset in $Y$, which is often $\mathbb{R}^k$ |
| $T$ | subset in $Z$, which is often $\mathbb{R}^l$ |
| $\mathcal{X}$ | enumerated set of all samples $x_1, \dots, x_i, \dots, x_N$ |
| $X$ | dataset of samples $x_{i_1}, \dots, x_{i_n}$ from $\mathcal{X}$ |
| $x$ | sample |
| $RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G}))$ | random variables from $(\Omega, \mathcal{F}, P)$ to $(Y, \mathcal{G})$ |

## 1. Introduction

Let $\mathcal{X} = \{x_1, \dots, x_i, \dots, x_N\}$ be a nonempty finite enumerated set of samples and let $X = \{x_{i_1}, \dots, x_{i_n}\}$ be a finite dataset of possibly repeated samples from $\mathcal{X}$. There is a bijection between the set of all such datasets and $\mathbb{N}^{|\mathcal{X}|}$

$$\begin{aligned} \{\text{all datasets } X\} &\xrightarrow{\phi} \mathbb{N}^{|\mathcal{X}|} \\ X &\mapsto (n_1, \dots, n_i, \dots, n_N) \end{aligned}$$

where each entry $n_i$ is the number of times $x_i$ appears in $X$. This bijection lets us use $\mathbb{N}^{|\mathcal{X}|}$ as a convenient way to represent all datasets $X$.

**Example 1.1.** If $\mathcal{X} = \{x_1, \dots, x_{10}\}$ and $X = \{x_1, x_2, x_2\}$ then $X = (1, 2, 0, \dots, 0) \in \mathbb{N}^{10}$.

We define the difference between two datasets.

**Definition 1.2.** (difference) Write $X = (n_1, \dots, n_N)$ and $X' = (n'_1, \dots, n'_N)$. We define their difference $X - X' = (|n_1 - n'_1|, \dots, |n_N - n'_N|)$.

Surely this difference is symmetric $X - X' = X' - X$.

**Example 1.3.** (difference by one sample) We represent $X = \{x_1\} = (1, 0, \dots, 0)$ and $X' = \{x_1, x_2\} = (1, 1, 0, \dots, 0)$. Their difference is $X - X' = (0, 1, 0, \dots, 0) = \{x_2\}$.

**Example 1.4.** (difference by two samples) We represent $X = \{x_1\} = (1, 0, \dots, 0)$ and $X' = \{x_2\} = (0, 1, 0, \dots, 0)$. Their difference is $X - X' = (1, 1, 0, \dots, 0) = \{x_1, x_2\}$.

Let $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} Y, X \mapsto f(X)$ be a function that provides some information about each dataset $X$. Such $f$ is sometimes called a query in literature. In 2003, Nissim and Dinur showed in [3] that it is impossible to publish arbitrary $f(X)$ without revealing some information about some $x \in X$, and that all information about all $x$ can be revealed by publishing $f(X_j), j = 1, \dots, o$ for an $o$ far smaller than was implied by previous work. We state this problem.

**Problem**: how to publish plaintext $f(X), f(X')$ and still provide privacy for $X - X'$. That is, how to let someone know $f(X), f(X')$ and the absence/presence of $X - X'$ and not let that person learn the samples in $X - X'$.

In 2006, Dwork, McSherry, Nissim and Smith defined differential privacy as a solution and provided mechanisms to achieve differential privacy in [5]. It provides privacy for the difference $X - X'$, hence the name *differential privacy*. We state this solution.

**Solution**: replace $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} Y, X \mapsto f(X)$ with $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G})), X \mapsto M(X)$ such that

1. (accuracy) $M(X)$ is concentrated around $f(X)$ so that we can publish samples from $M(X)$ in place of $f(X)$
2. (privacy) $M(X), M(X')$ are close in terms of the difference $X - X'$ to provide privacy for the samples in $X - X'$

This solution comes from the intuition

- samples' privacy can not be compromised by any $f(X)$ if they are not in $X$
- therefore let them have roughly the same privacy when they are in $X$ as the privacy they have when they are not in $X$

The distance between samples from $M(X)$ and $f(X)$ is our accuracy loss. The proximity between $M(X)$ and $M(X')$ is our privacy gain. We provide details about above problem and solution in the rest of this paper. It is worth noting that this solution has since been either generalized or modified or relaxed to other notions such as Rényi differential privacy, mean-concentrated differential privacy, zero-concentrated differential privacy, among others. We do not discuss these notions in this paper.

**Example 1.5.** Professor published midterm 1 average score $s$, when Mike had not taken the test. He publishes midterm 2 average score $s'$, when Mike did take the test. If

$$s \not\approx s'$$

then we infer about Mike's score. Or if

$$P(s \text{ is high}) \not\approx P(s' \text{ is high})$$

then we infer about Mike's score.

**Example 1.6.** Company published average salary $\mu$ before ABC joined. Company publish average salary $\mu'$ after ABC joined. If

$$\mu \not\approx \mu'$$

then we infer about ABC's salary. Or if

$$P(\mu \text{ is in any range}) \approx P(\mu' \text{ is in any range})$$

then we infer not about ABC's salary.

**Example 1.7.** Researcher published model parameters $\theta$ trained on lung images before yours was added. Researcher publishes model parameters $\theta'$ trained on lung images after yours was added. If

$$\theta \approx \theta'$$

then we infer not about your lung image. Or if

$$P(\theta \text{ is in any range}) \approx P(\theta' \text{ is in any range})$$

then we infer not about your lung image.

## 2. Differential Privacy

We recall some definitions for datasets.

**Definition 2.1.** We define 1-norm for datasets as $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{||-||_1} \mathbb{R}, X \mapsto \sum_{i=1}^{|\mathcal{X}|} |n_i|$.

This 1-norm counts the number of samples in each dataset. Its induced distance counts the number of samples in the difference between two datasets.

**Definition 2.2.** We define 1-distance between two datasets as $\mathbb{N}^{|\mathcal{X}|} \times \mathbb{N}^{|\mathcal{X}|} \xrightarrow{d_1} \mathbb{R}, (X, X') \mapsto ||X - X'||_1$.

**Example 2.3.** The distance between $X, X'$ in example 1.3 is $d_1(X, X') = ||(1, 0, \ldots, 0)||_1 = 1$.

**Example 2.4.** The distance between $X, X'$ in example 1.4 is $d_1(X, X') = ||(1, 1, 0, \ldots, 0)||_1 = 2$.

We will use the following two lemmas about 1-norm and 1-distance for datasets with respect to partition in subsection 2.4.

**Lemma 2.5.** If $\mathcal{X} = \bigsqcup_{j=1}^{o} \mathcal{X}_j$ is a partition of $\mathcal{X}$ then $|X|_1 = \sum_{j=1}^{o} |X \cap \mathcal{X}_j|_1$.

*Proof.* This follows from the fact that 1-norm counts the number of samples in $X$. □

**Lemma 2.6.** If $\mathcal{X} = \bigsqcup_{j=1}^{o} \mathcal{X}_j$ is a partition of $\mathcal{X}$ then $d_1(X, X') = \sum_{j=1}^{o} d_1(X \cap \mathcal{X}_j, X' \cap \mathcal{X}_j)$.

*Proof.* This follows from the fact that 1-distance counts the number of samples in $X - X'$ and $X - X' = \bigsqcup_{j=1}^{o} ((X - X') \cap \mathcal{X}_j = \bigsqcup_{j=1}^{o} ((X \cap \mathcal{X}_j) - (X' \cap \mathcal{X}_j))$. □

We also recall some definitions for functions.

**Definition 2.7.** A function $(X, d_X) \xrightarrow{f} (Y, d_Y)$ between two metric spaces is called Lipschitz continuous if there exists a real constant $c$ such that $d_Y(f(x), f(x')) \leq c d_X(x, x')$ for all $x, x' \in X$. Such a constant $c$ is called a Lipschitz constant, and the smallest one is called the Lipschitz constant and denoted by $\kappa$.

We modify this definition of Lipschitz continuity and Lipschitz constant for use in subsection 3.3.

**Definition 2.8.** A function $(X \times Y, d_X) \xrightarrow{f} (Z, d_Z)$ is called modified Lipschitz continuous if there exists a real constant $c$ such that $d_Z(f(x, y), f(x', y)) \leq c d_X(x, x')$ for all $x, x' \in X, y \in Y$. Such a constant $c$ is called a modified Lipschitz constant, and the smallest one is called the modified Lipschitz constant and denoted by $\kappa$.

We also add a definition of Lipschitz constant at each point for use in subsection 3.5.

**Definition 2.9.** A constant $c$ is called a Lipschitz constant for $x$ if $d_Y(f(x), f(x')) \leq c d_X(x, x')$ for all $x' \in X$. The smallest such $c$ is called the Lipschitz constant for $x$ and denoted by $\kappa(x)$.

We assume that $\kappa(x)$ exists for all $x \in X$. Surely $\kappa = \sup\{\kappa(x), x \in X\}$ and $\kappa(x) \leq \kappa$ for all $x \in X$.

**Example 2.10.** For identity function $(Y,d) \xrightarrow{\text{id}} (Y,d), y \mapsto y$ we have $d(\text{id}(y), \text{id}(y')) = d(y,y')$ for all $y, y' \in Y$. Hence $\kappa = 1$. Also $\kappa(x) = 1$ for all $x \in X$.

**Example 2.11.** For this function

$$\begin{aligned}(\{0,1,2,3\}, d_1) &\xrightarrow{f} (\{0,1,2,5\}, d_1)\\ 0 &\mapsto 0\\ 1 &\mapsto 1\\ 2 &\mapsto 2\\ 3 &\mapsto 5\end{aligned}$$

we have

$$\begin{aligned}|f(0) - f(0)| &= 0|0-0|\\ |f(0) - f(1)| &= 1|0-1|\\ |f(0) - f(2)| &= 1|0-2|\\ |f(0) - f(3)| &= \frac{5}{3}|0-3|\end{aligned}$$

Hence $\kappa(0) = \frac{5}{3}$. Similarly $\kappa(1) = 2, \kappa(2) = 3, \kappa(3) = 3$ while $\kappa = 3$.

**Example 2.12.** For binary function $\mathcal{X} \xrightarrow{b} \{0,1\}, x \mapsto b(x)$ and counting function $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}, d_1), X \mapsto \sum\limits_{x \in X} b(x)$ we have

$$\begin{aligned}d_1(f(X), f(X')) &= |f(X) - f(X')|\\ &= |\sum_{x \in X} b(x) - \sum_{x' \in X'} b(x')|\\ &\leq \sum_{x \in X - X'} b(x)\\ &\leq \sum_{x \in X - X'} 1\\ &= d_1(X - X')\end{aligned}$$

for all datasets $X, X'$. Hence $\kappa \leq 1$. If there are datasets $X, X'$ that differ by some samples $x_{i_1}, \ldots, x_{i_k}$ and $b(x_{i_f}) = 1$ then $\kappa = 1$.

**Example 2.13.** For binning function $\mathcal{X} \xrightarrow{b} \{0,1\}^k, x_i \mapsto (0,\dots 0,1,0,\dots,0)$ and histogram function $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}^k, d_1), X \mapsto \sum_{i=1}^{n} b(x_i)$ we have

$$\begin{aligned} d_1(f(X), f(X')) &= ||f(X) - f(X')||_1 \\ &= ||\sum_{x\in X} b(x) - \sum_{x'\in X'} b(x')||_1 \\ &\leq ||\sum_{x\in X-X'} b(x)||_1 \\ &= \sum_{x\in X-X'} ||b(x)||_1 \\ &= \sum_{x\in X-X'} 1 \\ &= d_1(X, X') \end{aligned}$$

for all datasets $X, X'$. Hence $\kappa \leq 1$. If there are datasets $X, X'$ that differ by some samples $x_{i_1}, \dots, x_{i_k}$ then $\kappa = 1$.

Now we carry out the solution of differential privacy.

**Definition 2.14.** A random mechanism is a map $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G})), X \mapsto M(X)$.

**Definition 2.15.** A random mechanism $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G})), X \mapsto M(X)$ is called $(\epsilon, \delta)$-differentially private if

$$P(M(X)^{-1}(S)) \leq e^{d_1(X,X')\epsilon} P(M(X')^{-1}(S)) + \delta$$

for all datasets $X, X'$ and all subsets $S \in \mathcal{G}$.

**Definition 2.16.** A random mechanism $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G})), X \mapsto M(X)$ is called $\epsilon$-differentially private if it is $(\epsilon, 0)$-differentially private.

Definition 2.15 is sometimes called $\delta$-approximate differential privacy while definition 2.16 is sometimes called pure differential privacy in literature. Furthermore $\epsilon$ is called privacy cost while $\delta$ is not called any name. We note that

- sample space $\Omega$, $\sigma$-algebra $\mathcal{F}$ and probability measure $P$ are not made explicit
- definition 2.15 means
  - $\ln(P(M(X)^{-1}(S)) - \delta) - \ln(P(M(X')^{-1}(S))) \leq d_1(X, X')\epsilon$
  - $\ln(P(M(X')^{-1}(S)) - \delta) - \ln(P(M(X)^{-1}(S))) \leq d_1(X, X')\epsilon$
- definition 2.16 means
  - $|\ln(P(M(X)^{-1}(S))) - \ln(P(M(X')^{-1}(S)))| \leq d_1(X, X')\epsilon$
- smaller $\epsilon$ means better privacy
- todo: interpretation for $\delta$
- while the usual definitions of differential privacy work with datasets of distance 1, definitions 2.15 and 2.16 work with datasets of arbitrary distance, thus removing the need to discuss group privacy in [4, 2.3.2].

We provide another formulation of $(\epsilon, \delta)$-differential privacy and $\epsilon$-differential privacy.

**Definition 2.17.** For a random variable $(\Omega, \mathcal{F}, P) \xrightarrow{V} (Y, \mathcal{G})$, we define its support $\text{supp}(V)$ to be the smallest closed subset $S^\circ \in \mathcal{G}$ such that $P(V^{-1}(S^\circ)) = 1$.

We assume that $\text{supp}(V)$ exists for every random variable $V$.

**Definition 2.18.** For two random variables $(\Omega, \mathcal{F}, P) \xrightarrow[V']{V} (Y, \mathcal{G})$, we define their $\delta$-approximate max divergence $D_\infty^\delta(V||V') = \max\{\ln \frac{P(V^{-1}(S))-\delta}{P(V'^{-1}(S))}, S \in \mathcal{G}, S \subset \text{supp}(V) \text{ and } P(V^{-1}(S)) \geq \delta\}$.

This definition is equivalent to the definition $D_\infty^\delta(V||V') = \max\{\ln \frac{P(V^{-1}(S))-\delta}{P(V'^{-1}(S))}, S \in \mathcal{G} \text{ such that } P(V^{-1}(S) \geq \delta\}$ by definition of $\text{supp}(V)$.

**Definition 2.19.** For two random variables $(\Omega, \mathcal{F}, P) \xrightarrow[V']{V} (Y, \mathcal{G})$, we define their max divergence $D_\infty(V||V') = D_\infty^0(V||V')$.

Compare this divergence with Kullback-Leibler divergence. Now we see

- a random mechanism $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G}))$ is $(\epsilon, \delta)$-differentially private iff $D_\infty^\delta(M(X)||M(X')) \leq d_1(X, X')\epsilon$ and $D_\infty^\delta(M(X')||M(X)) \leq d_1(X, X')\epsilon$ for all datasets $X, X'$.
- a random mechanism $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G}))$ is $\epsilon$-differentially private iff $D_\infty(M(X)||M(X')) \leq d_1(X, X')\epsilon$ and $D_\infty(M(X')||M(X)) \leq d_1(X, X')\epsilon$ for all datasets $X, X'$.

**Example 2.20.** (perfect privacy) If $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G}))$ is a 0-differentially private random mechanism then $P(M(X)^{-1}(S)) = P(M(X')^{-1}(X))$ for all datasets $X, X'$ and all subsets $S \in \mathcal{G}$. The difference $X - X'$ makes no difference in the outcomes $P(M(X)^{-1}(S)), P(M(X')^{-1}(S))$. See the next three examples for what this perfect privacy may cause to accuracy.

**Example 2.21.** (perfect privacy, perfect accuracy) If we replace the zero function

$$\begin{aligned} \{0,1\} &\xrightarrow{0} \{0,1\} \\ 0 &\mapsto 0 \\ 1 &\mapsto 0 \end{aligned}$$

with

$$\begin{aligned} \{0,1\} &\xrightarrow{M} RV((\Omega, \mathcal{F}, P), (\{0,1\}, \mathcal{G})) \\ 0 &\mapsto 0 \\ 1 &\mapsto 0 \end{aligned}$$

then

- $P(M(0) = y) = P(M(1) = y)$ for all $y \in \{0, 1\}$ (perfect privacy)

- $P(M(x) = f(x)) = 1$ for all $x \in \{0, 1\}$ (perfect accuracy)

**Example 2.22.** (perfect privacy, partial accuracy) If we replace the identity function

$$\begin{aligned} \{0,1\} &\xrightarrow{\text{id}} \{0,1\} \\ 0 &\mapsto 0 \\ 1 &\mapsto 1 \end{aligned}$$

with

$$\begin{aligned} \{0,1\} &\xrightarrow{M} RV((\Omega, \mathcal{F}, P), (\{0,1\}, \mathcal{G})) \\ 0 &\mapsto 0 \\ 1 &\mapsto 0 \end{aligned}$$

then

- $P(M(0) = y) = P(M(1) = y)$ for all $y \in \mathbb{R}$ (perfect privacy)
- $P(M(0) = f(0)) = 1$ while $P(M(1) = f(1)) = 0$ (partial accuracy)

**Example 2.23.** (perfect privacy, zero accuracy) If we replace the identity function

$$\begin{aligned} \{0,1\} &\xrightarrow{\text{id}} \{0,1\} \\ 0 &\mapsto 0 \\ 1 &\mapsto 1 \end{aligned}$$

with

$$\begin{aligned} \{0,1\} &\xrightarrow{M} RV((\Omega, \mathcal{F}, P), (\{0,1\}, \mathcal{G})) \\ 0 &\mapsto 2 \\ 1 &\mapsto 2 \end{aligned}$$

then

- $P(M(0) = y) = P(M(1) = y)$ for all $y \in \mathbb{R}$ (perfect privacy)
- $P(M(0) = f(0)) = 0$ and $P(M(1) = f(1)) = 0$ (zero accuracy)

We will use the following lemmas about $\delta$-approximate max divergence in subsection 2.4.

**Lemma 2.24.** For two joint random variables $(\Omega, \mathcal{F}, P) \xrightarrow[(V_1', \ldots, V_o')]{(V_1, \ldots, V_o)} (Y, \mathcal{G})$, if the $V_j$ are independent and the $V_j'$ are independent then $D_\infty^\delta((V_1, \ldots, V_o)||(V_1', \ldots, V_o')) = \sum_{j=1}^{o} D_\infty^\delta(V_j||V_j')$.

*Proof.* todo: verify equality, whether independence is needed, whether inequality is wanted instead of equality. □

**Lemma 2.25.** For two random variables $(\Omega, \mathcal{F}, P) \underset{V'}{\overset{V}{\rightrightarrows}} (Y, \mathcal{G})$, if $\delta \leq \delta'$ then $D_\infty^{\delta'}(V||V') \leq D_\infty^{\delta}(V||V')$.

*Proof.* By definition

$$\begin{aligned} D_\infty^{\delta'}(V||V') &= \max\{\ln \frac{P(V^{-1}(S)) - \delta'}{P(V'^{-1}(S))}, S \in \mathcal{G} \text{ such that } P(V^{-1}(S) \geq \delta'\} \\ &\leq \max\{\ln \frac{P(V^{-1}(S)) - \delta}{P(V'^{-1}(S))}, S \in \mathcal{G} \text{ such that } P(V^{-1}(S) \geq \delta\} \\ &= D_\infty^{\delta}(V||V') \end{aligned}$$

□

Beside quantifying accuracy loss and privacy gain in 1 and 2, differential privacy has the following nice properties.

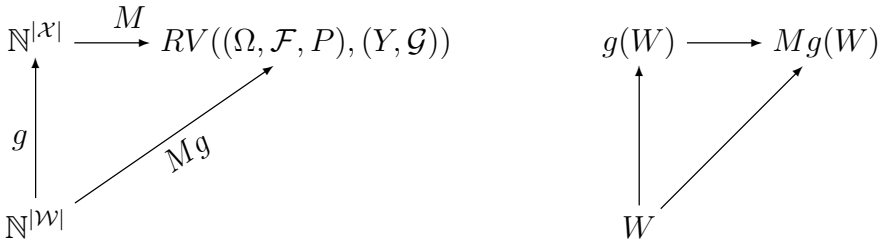
2.1. **Preprocessing.** One can preprocess the datasets before using an $(\epsilon, \delta)$-differentially private random mechanism and the resulted random mechanism is less differentially private by a factor of Lipschitz constant.

**Proposition 2.26.** (Preprocessing) If $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G}))$ is $(\epsilon, \delta)$-differentially private and $\mathbb{N}^{|\mathcal{W}|} \xrightarrow{g} \mathbb{N}^{|\mathcal{X}|}$ is any preprocessing with Lipschitz constant $\kappa$ then the modification $\mathbb{N}^{|\mathcal{W}|} \xrightarrow{Mg} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G}))$ is $(\kappa\epsilon, \delta)$-differentially private.

*Proof.* By definition

$$\begin{aligned} P((Mg)(W)^{-1}(S)) &= P(M(g(W))^{-1}(S)) \\ &\leq e^{d_1(g(W), g(W'))\epsilon} P(M(g(W')^{-1}(S)) + \delta \\ &\leq e^{\kappa d_1(W, W')\epsilon} P((Mg)(W')^{-1}(S)) + \delta \end{aligned}$$

for all datasets $W, W'$ and subsets $S \in \mathcal{G}$.




□

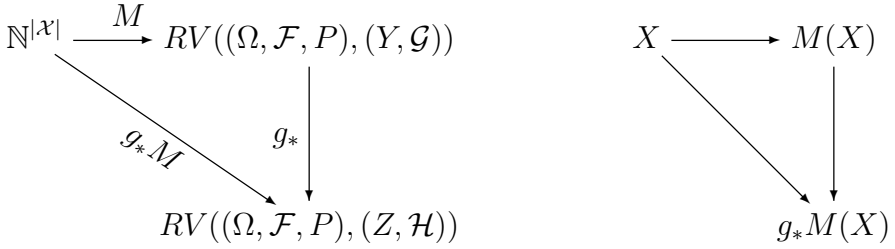
2.2. **Postprocessing.** One can postprocess the outputs of $M(X)$ of an $(\epsilon, \delta)$-differentially private random mechanism $M$ and the resulted random mechanism is equally differentially private. This also means an adversary can not modify the outputs of $M(X)$ to make $M$ less differentially private.

**Proposition 2.27.** (Postprocessing) If $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G}))$ is $(\epsilon, \delta)$-differentially private and $(Y, \mathcal{G}) \xrightarrow{g} (Z, \mathcal{H})$ is any measurable postprocessing then the modification $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{g_*M} RV((\Omega, \mathcal{F}, P), (Z, \mathcal{H}))$ is $(\epsilon, \delta)$-differentially private, where $RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G})) \xrightarrow{g_*} RV((\Omega, \mathcal{F}, P), (Z, \mathcal{H})), V \mapsto gV$ is the pushforward of $g$.

*Proof.* By definition

$$\begin{aligned} P((g_*M)(X)^{-1}(T)) &= P(g_*(M(X))^{-1}(T)) \\ &= P((gM(X))^{-1}(T)) \\ &= P(M(X)^{-1}(g^{-1}(T))) \\ &\leq e^{d_1(X,X')\epsilon} P(M(X')^{-1}(g^{-1}(T))) + \delta \\ &= e^{d_1(X,X')\epsilon} P((g_*M)(X')^{-1}(T)) + \delta \end{aligned}$$

for all datasets $X, X'$ and subsets $T \in \mathcal{H}$. The same holds for $P((g_*M)(X')^{-1}(T))$.




□

2.3. **Sequential Composition.** Any sequence of differentially private mechanism $M_1, \ldots, M_m$ is also differentially private. This is true not only when they provide independent information $M_1(X), \ldots, M_m(X)$ but also when $M_h(X)$ incorporates $M_1(X), \ldots, M_{h-1}(X)$.

**Proposition 2.28.** If $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M_h} RV((\Omega, \mathcal{F}, P), (Y_h, \mathcal{G}_h)), 1 \leq h \leq m$ are independent $(\epsilon_h, \delta_h)$-differentially private then their sequential composition $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{(M_1,\ldots,M_m)} RV((\Omega, \mathcal{F}, P), (\prod_{h=1}^{m} Y_h, \mathcal{G}(\prod_{h=1}^{m} Y_h))), X \mapsto (M_1(X), \ldots, M_m(X))$ is $(\sum_{h=1}^{m} \epsilon_h, \sum_{h=1}^{m} \delta_h)$-differentially private.

*Proof.* Write $\epsilon = \sum_{h=1}^{m} \epsilon_h$ and $\delta = \sum_{h=1}^{m} \delta_h$. Then

$$\begin{aligned} D_\infty^\delta(M_1(X), \dots, M_m(X) || M_1(X'), \dots, M_m(X')) &= \sum_{h=1}^{m} D_\infty^\delta(M_h(X)||M_h(X')) \text{ by lemma 2.24} \\ &\leq \sum_{h=1}^{m} D_\infty^{\delta_h}(M_h(X)||M_h(X')) \text{ by lemma 2.25} \\ &\leq \sum_{h=1}^{m} d_1(X, X')\epsilon_h \\ &= d_1(X, X')\epsilon \end{aligned}$$

for all datasets $X, X'$. The same holds for $D_\infty^\delta(M_1(X'), \dots, M_m(X')||M_1(X), \dots, M_m(X))$. □

**Corollary 2.29.** If $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M_h} RV((\Omega, \mathcal{F}, P), (Y_h, \mathcal{G}_h)), 1 \leq h \leq m$ are $\epsilon_h$-differentially private then their sequential composition $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{(M_1, \dots, M_m)} RV((\Omega, \mathcal{F}, P), (\prod_{h=1}^{m} Y_h, \mathcal{G}(\prod_{h=1}^{m} Y_h))), X \mapsto (M_1(X), \dots, M_m(X))$ is $\sum_{h=1}^{m} \epsilon_h$-differentially private.

*Proof.* This follows immediately from proposition 2.28. □

**Claim 2.30.** (other) If $\mathbb{P}(\mathcal{X}) \xrightarrow{g_i} \mathbb{R}^l, 1 \leq h \leq l$ are independent $(\epsilon, \delta)$-differentially private randomized algorithms then their composition $\mathbb{P}(\mathcal{X}) \xrightarrow{(g_1, \dots, g_k)} \mathbb{R}^{kl}$ is $(\sqrt{k}\epsilon \ln \frac{1}{\delta'}, k\delta + \delta')$-differentially private for $\delta' > 0$.

*Proof.* □

**Claim 2.31.** (advanced) If $\mathbb{P}(\mathcal{X}) \xrightarrow{g_i} \mathbb{R}^l, 1 \leq h \leq l$ are independent $(\epsilon, \delta)$-differentially private randomized algorithms then their adaptive composition $\mathbb{P}(\mathcal{X}) \xrightarrow{(g_1, \dots, g_k)} \mathbb{R}^{kl}$ is $(\sqrt{2k \ln \frac{1}{\delta'}}\epsilon + k\epsilon(e^\epsilon - 1), k\delta + \delta')$-differentially private for $\delta' > 0$.

*Proof.* □

todo: reconcile above 3 claims.

2.4. **Parallel Composition.** While the privacy cost of sequential composition $(M_1, \dots, M_m)$ of differentially private mechanisms adds up, the privacy cost of parallel composition $(M_1(- \cap \mathcal{X}_{j_1}), \dots, M_m(- \cap \mathcal{X}_{j_m}))$ does not, where $\mathcal{X} = \bigsqcup_{j=1}^{o} \mathcal{X}_j$ is a partition of $\mathcal{X}$ and $j_1, \dots, j_m$ are unique. This is useful when we can replace $M$ or $(M_1, \dots, M_m)$ with $(M_1(- \cap \mathcal{X}_{j_1}), \dots, M_m(- \cap \mathcal{X}_{j_m}))$.

**Proposition 2.32.** Let $\mathcal{X} = \bigsqcup_{j=1}^{o} \mathcal{X}_j$ be a partition of $\mathcal{X}$ and let $j_1, \dots, j_m$ be unique. If $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M_h} RV((\Omega, \mathcal{F}, P), (Y_h, \mathcal{G}_h)), 1 \leq h \leq m$ are $(\epsilon_h, \delta_h)$-differentially private then their

parallel composition
$\mathbb{N}^{|\mathcal{X}|} \xrightarrow{(M_1(-\cap\mathcal{X}_{j_1}),\ldots,M_m(-\cap\mathcal{X}_{j_m}))} RV((\Omega,\mathcal{F},P),(\prod_{h=1}^{m} Y_h, \mathcal{G}(\prod_{h=1}^{m} Y_h))), X \mapsto (M_1(X\cap\mathcal{X}_{j_1}),\ldots,M_m(X\cap\mathcal{X}_{j_m}))$ is $(\max\{\epsilon_h, 1\leq h\leq m\}, \max\{\delta_h, 1\leq h\leq m\})$-differentially private.

*Proof.* Write $X_h = X\cap\mathcal{X}_{j_h}, X'_h = X'\cap\mathcal{X}_{j_h}$, $\epsilon = \max\{\epsilon_h, 1\leq h\leq m\}$ and $\delta = \max\{\delta_h, 1\leq h\leq m\}$. Then

$$\begin{aligned}
D_\infty^\delta(M_1(X_1),\ldots,M_m(X_m)||M_1(X'_1),\ldots,M_m(X'_m)) &= \sum_{h=1}^{m} D_\infty^\delta(M_h(X_i)||M_h(X'_i)) \text{ by lemma 2.24}\\
&\leq \sum_{h=1}^{m} D_\infty^{\delta_h}(M_h(X_i)||M_h(X'_i)) \text{ by lemma 2.25}\\
&\leq \sum_{h=1}^{m} d_1(X_h,X'_h)\epsilon_h\\
&\leq \sum_{h=1}^{m} d_1(X_h,X'_h)\epsilon\\
&\leq \sum_{j=1}^{o} d_1(X_j,X'_j)\epsilon\\
&= d_1(X,X')\epsilon \text{ by lemma 2.6}
\end{aligned}$$

for all datasets $X, X'$. The same holds for $D_\infty^\delta(M_1(X'_1),\ldots,M_m(X'_m)||M_1(X_1),\ldots,M_m(X_m))$. □

**Corollary 2.33.** Let $\mathcal{X} = \bigsqcup_{j=1}^{o} \mathcal{X}_j$ be a partition of $\mathcal{X}$ and let $j_1,\ldots,j_m$ be unique. If $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M_h} RV((\Omega,\mathcal{F},P),(Y_h,\mathcal{G}_h)), 1\leq h\leq m$ are $\epsilon_h$-differentially private then their parallel composition
$\mathbb{N}^{|\mathcal{X}|} \xrightarrow{(M_1(-\cap\mathcal{X}_{j_1}),\ldots,M_m(-\cap\mathcal{X}_{j_m}))} RV((\Omega,\mathcal{F},P),(\prod_{h=1}^{m} Y_h, \mathcal{G}(\prod_{h=1}^{m} Y_h))), X \mapsto (M_1(X\cap\mathcal{X}_{j_1}),\ldots,M_m(X\cap\mathcal{X}_{j_m}))$ is $\max\{\epsilon_h, 1\leq h\leq m\}$-differentially private.

*Proof.* This follows immediately from proposition 2.33. □

## 3. Mechanisms for Differential Privacy

Given deterministic

$$\begin{aligned}
\mathbb{N}^{|\mathcal{X}|} &\xrightarrow{f} \mathbb{R}^k\\
X &\mapsto f(X)
\end{aligned}$$

we create random mechanisms

$$\begin{aligned}
\mathbb{N}^{|\mathcal{X}|} &\xrightarrow{M} RV((\Omega,\mathcal{F},P),(Y,\mathcal{G}))\\
X &\mapsto M(X)
\end{aligned}$$

such that 1 and 2 hold.

3.1. **Randomized Response Mechanism.** We use Bernoulli random variable to create random mechanisms such that 1 and 2 hold.

**Definition 3.1.** For $p, q \in (0,1)$ and sequences of response $\{0,1\}^k \xrightarrow{\text{id}} \{0,1\}^k$, we define its randomized response mechanism

$$\begin{aligned} \{0,1\}^k &\xrightarrow{M_{\text{id},B(p),B(q)}} RV((\Omega,\mathcal{F},P),(\{0,1\}^k,\mathcal{G})) \\ (\dots,b_f,\dots) &\mapsto M_{\text{id},B(p),B(q)}(\dots,b_f,\dots) \end{aligned}$$

where $(\Omega,\mathcal{F}) \xrightarrow{M_{\text{id},B(p),B(q)}(\dots,b_f,\dots)} (\{0,1\}^k,\mathcal{G})$ is the random variable defined by

$$\begin{aligned} P(M_{\text{id},B(p),B(q)}(\dots,0,\dots) = (\dots,0,\dots)) &= 1-p \\ P(M_{\text{id},B(p),B(q)}(\dots,0,\dots) = (\dots,1,\dots)) &= p \\ P(M_{\text{id},B(p),B(q)}(\dots,1,\dots) = (\dots,0,\dots)) &= q \\ P(M_{\text{id},B(p),B(q)}(\dots,1,\dots) = (\dots,1,\dots)) &= 1-q \end{aligned}$$

**Proposition 3.2.** For sequences of response $(\{0,1\}^k,d_1) \xrightarrow{\text{id}} (\{0,1\}^k,d_1)$, if $\max\left\{\frac{1-p}{q},\frac{p}{1-q},\frac{q}{1-p},\frac{1-q}{p}\right\} \leq e^\epsilon$ then its randomized response mechanism $M_{\text{id},B(p),B(q)}$ is $\epsilon$-differentially private.

*Proof.* Write $c = \max\left\{\frac{1-p}{q},\frac{p}{1-q},\frac{q}{1-p},\frac{1-q}{p}\right\}$. Then

$$\begin{aligned} \frac{P(M_{\text{id},B(p),B(q)}(\dots,b_f,\dots) = (\dots,y_f,\dots))}{P(M_{\text{id},B(p),B(q)}(\dots,b'_f,\dots) = (\dots,y_f,\dots))} &= \frac{\prod_{b_f=0,y_f=0} 1-p \prod_{b_f=0,y_f=1} p \prod_{b_f=1,y_f=0} q \prod_{b_f=1,y_f=1} 1-q}{\prod_{b'_f=0,y_f=0} 1-p \prod_{b'_f=0,y_f=1} p \prod_{b'_f=1,y_f=0} q \prod_{b'_f=1,y_f=1} 1-q} \\ &\leq \prod_{b_f \neq b'_f} c \\ &= c^{d_1((\dots,b_f,\dots),(\dots,b'_f,\dots))} \\ &\leq e^{d_1((\dots,b_f,\dots),(\dots,b'_f,\dots))\epsilon} \end{aligned}$$

for all sequences $(\dots,b_f,\dots),(\dots,b'_f,\dots)$. Similarly $\frac{P(M_{\text{id},B(p),B(q)}(\dots,b'_f,\dots)=(\dots,y_f,\dots))}{P(M_{\text{id},B(p),B(q)}(\dots,b_f,\dots)=(\dots,y_f,\dots))} \leq e^{d_1((\dots,b_f,\dots),(\dots,b'_f,\dots))\epsilon}$. Hence $|\ln(P(M_{\text{id},B(p),B(q)}(\dots,b_f,\dots) = (\dots,y_f,\dots))) - \ln(P(M_{\text{id},B(p),B(q)}(\dots,b'_f,\dots) = (\dots,y_f,\dots)))| \leq d_1((\dots,b_f,\dots),(\dots,b'_f,\dots))\epsilon$ as desired. □

We compute how well $M_{\text{id},B(p),B(q)}(\dots,b_f,\dots)$ is concentrated around $(\dots,b_f,\dots)$.

**Proposition 3.3.** For $0 \leq j \leq k$ and sequences of response $\{0,1\}^k \xrightarrow{\text{id}} \{0,1\}^k$, its randomized response mechanism $M_{\text{id},B(p),B(q)}$ satisfies

$$P(||M_{\text{id},B(p),B(q)}(\dots,b_f,\dots)-(\dots,b_f,\dots)||_1 = j) \leq C(k,j)\max\{p,q\}^j \max\{1-p,1-q\}^{k-j}$$

for every sequence $(\dots,b_f,\dots)$.

*Proof.* Surely

$$P(||M_{\mathrm{id},B(p),B(q)}(\dots,b_f,\dots)-(\dots,b_f,\dots)||_1=j) = P(j \text{ bits get flipped and } (k-j) \text{ bits stay put}) \\ \leq C(k,j)\max\{p,q\}^j \max\{1-p,1-q\}^{k-j}$$

□

**Corollary 3.4.** For $0 \leq j \leq k$ and sequences of response $\{0,1\}^k \xrightarrow{\mathrm{id}} \{0,1\}^k$, if $\max\left\{\frac{1-p}{q}, \frac{p}{1-q}, \frac{q}{1-p}, \frac{1-q}{p}\right\} \leq e^\epsilon$ then its randomized response mechanism $M_{\mathrm{id},B(p),B(q)}$ is $\epsilon$-differentially private and

$$P(||M_{\mathrm{id},B(p),B(q)}(\dots,b_f,\dots)-(\dots,b_f,\dots)||_1=j) \leq C(k,j)\max\{p,q\}^j \max\{1-p,1-q\}^{k-j}$$

for every sequence $(\dots,b_f,\dots)$.

*Proof.* This follows immediately from proposition 3.2 and proposition 3.3. □

See example 4.1 for an illustration of the randomized response mechanism. It can also be generalized for sequences of response $\{0,\dots,d\}^k \xrightarrow{\mathrm{id}} \{0,\dots,d\}^k$.

3.2. **Laplace Mechanism.** We can use Laplace random variable to create random mechanisms such that 1 and 2 hold.

**Definition 3.5.** For $\beta > 0$ and $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} \mathbb{R}^k$, we define its Laplace mechanism

$$\begin{aligned} \mathbb{N}^{|\mathcal{X}|} &\xrightarrow{M_{f,L(\beta)}} RV((\Omega,\mathcal{F},P),(\mathbb{R}^k,\mathcal{B}(\mathbb{R}^k))) \\ X &\mapsto f(X)+(L(\beta),\dots,L(\beta)) \end{aligned}$$

where $(\Omega,\mathcal{F}) \xrightarrow{L(\beta)} (\mathbb{R},\mathcal{B}(\mathbb{R}))$ is the Laplace random variable with probability density function $f_{L(\beta)}(x) = \frac{1}{2\beta}e^{-\frac{|x|}{\beta}}$.

**Proposition 3.6.** For $(\mathbb{N}^{|\mathcal{X}|},d_1) \xrightarrow{f} (\mathbb{R}^k,d_1)$ with Lipschitz constant $\kappa$, if $\beta \geq \frac{\kappa}{\epsilon}$ then its Laplace mechanism $M_{f,L(\beta)}$ is $\epsilon$-differentially private.

*Proof.* Write $f(X) = (y_1, \ldots, y_k)$ and $f(X') = (y'_1, \ldots, y'_k)$. Let $f_X$ and $f_{X'}$ denote the probability density functions of $M_{f,L(\beta)}(X)$ and $M_{f,L(\beta)}(X')$. Then

$$\begin{aligned}
\frac{f_X(z_1, \ldots, z_k)}{f_{X'}(z_1, \ldots, z_k)} &= \prod_{f=1}^{k} \frac{f_{L(\beta)}(z_f - y_f)}{f_{L(\beta)}(z_f - y'_f)} \\
&= \prod_{f=1}^{k} \frac{e^{-\frac{|z_f - y_f|}{\beta}}}{e^{-\frac{|z_f - y'_f|}{\beta}}} \\
&= \prod_{f=1}^{k} e^{\frac{|z_f - y'_f| - |z_f - y_f|}{\beta}} \\
&\leq \prod_{f=1}^{k} e^{\frac{|y_f - y'_f|}{\beta}} \\
&= e^{\frac{\sum_{f=1}^{k} |y_f - y'_f|}{\beta}} \\
&= e^{\frac{d_1(f(X), f(X)')}{\beta}} \\
&\leq e^{\frac{\kappa d_1(X, X')}{\beta}} \\
&\leq e^{d_1(X, X')\epsilon}
\end{aligned}$$

for all datasets $X, X'$. Similarly $\frac{f_{X'}(z_1,\ldots,z_k)}{f_X(z_1,\ldots,z_k)} \leq e^{d_1(X,X')\epsilon}$. Hence $|\ln(f_X(z_1, \ldots, z_k)) - \ln(f_{X'}(z_1, \ldots, z_k))| \leq d_1(X, X')\epsilon$ as desired. □

We compute how well $M_{f,L(\beta)}(X)$ is concentrated around $f(X)$.

**Proposition 3.7.** For $\delta \in (0, 1]$ and $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} \mathbb{R}^k$, its Laplace mechanism $M_{f,L(\beta)}$ satisfies

$$P\left(||M_{f,L(\beta)}(X) - f(X)||_\infty \geq \beta \ln\left(\frac{k}{\delta}\right)\right) \leq \delta$$

for every dataset $X$.

*Proof.* Surely

$$\begin{aligned}
P\left(||M_{f,L(\beta)}(X) - f(X)||_\infty \geq \beta \ln\left(\frac{k}{\delta}\right)\right) &= P\left(||f(X) + (L(\beta), \ldots, L(\beta)) - f(X)||_\infty \geq \beta \ln\left(\frac{k}{\delta}\right)\right) \\
&= P\left(||(L(\beta), \ldots, L(\beta))||_\infty \geq \beta \ln\left(\frac{k}{\delta}\right)\right) \\
&= P\left(\max\{|L_f(\beta)|, 1 \leq f \leq k\} \geq \beta \ln\left(\frac{k}{\delta}\right)\right) \\
&\leq kP\left(|L(\beta)| \geq \beta \ln\left(\frac{k}{\delta}\right)\right) \\
&= ke^{-\ln\left(\frac{k}{\delta}\right)} \\
&= \delta
\end{aligned}$$

□

**Corollary 3.8.** For $\delta \in (0, 1]$ and $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}^k, d_1)$ with Lipschitz constant $\kappa$, if $\beta = \frac{\kappa}{\epsilon}$ then its Laplace mechanism $M_{f,L(\beta)}$ is $\epsilon$-differentially private and satisfies

$$P\left(||M_{f,L(\beta)}(X) - f(X)||_\infty \geq \frac{\kappa}{\epsilon} \ln\left(\frac{k}{\delta}\right)\right) \leq \delta$$

for every dataset $X$.

*Proof.* This follows immediately from proposition 3.6 and proposition 3.7. □

See example 5.1 for an illustration of the Laplace mechanism.

3.3. **Exponential Mechanism.** Let $\mathbb{N}^{|\mathcal{X}|} \times Y \xrightarrow{s} \mathbb{R}, (X, y) \mapsto s(X, y)$ be a function such that for each $X$ the maximum $\max\{s(X, y), y \in Y\}$ exists and is realized $\max\{s(X, y), y \in Y\} = s(X, y_X)$ by some $y_X$. Denote the set of all such $y_X$ by $Y_X$.

**Example 3.9.** Let $\mathbb{N}^{|\mathcal{X}|}$ be the collection of test sets, $\mathbb{R}^k$ be the space of models and $\mathbb{N}^{|\mathcal{X}|} \times \mathbb{R}^k \xrightarrow{s} \mathbb{R}, (X, y) \mapsto s(X, y)$ be a score function. Then for each test set $X$ the maximum score $\max\{s(X, y),\ \text{model } y \in \mathbb{R}^k\}$ exists and is realized $\max\{s(X, y),\ \text{model } y \in \mathbb{R}^k\} = s(X, y_X)$ by some model $y_X$.

Now suppose we want to create random mechanism $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{M} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G})), X \mapsto M(X)$ such that

1. $s(X, M(X))$ is concentrated around $s(X, y_X)$ so that we can publish samples from $M(X)$ in place of $y_X$. Compare this with 1
2. same as 2

We can use exponential random variable to create random mechanisms such that 1 and 2 hold.

**Definition 3.10.** For $\lambda > 0$ and $\mathbb{N}^{|\mathcal{X}|} \times Y \xrightarrow{s} \mathbb{R}$, we define its exponential mechanism

$$\begin{aligned} \mathbb{N}^{|\mathcal{X}|} &\xrightarrow{M_{s,E(\lambda)}} RV((\Omega, \mathcal{F}, P), (Y, \mathcal{G})) \\ X &\mapsto M_{s,E(\lambda)}(X) \end{aligned}$$

where $(\Omega, \mathcal{F}) \xrightarrow{M_{s,E(\lambda)}(X)} (Y, \mathcal{G})$ is the random variable defined by $P(M_{s,E(\lambda)}(X) = y) \propto e^{\lambda s(X,y)}$.

**Proposition 3.11.** For $(\mathbb{N}^{|\mathcal{X}|} \times Y, d_1) \xrightarrow{s} (\mathbb{R}, d_1)$ with modified Lipschitz constant $\kappa$, if $\lambda \leq \frac{\epsilon}{2\kappa}$ then its exponential mechanism $M_{s,E(\lambda)}$ is $\epsilon$-differentially private.

*Proof.* For simplicity, assume $Y = \{y_1, \ldots, y_k\}$ finite. Then

$$\begin{aligned} \frac{P(M_{s,E(\lambda)}(X) = y)}{P(M_{s,E(\lambda)}(X') = y)} &= \frac{\frac{e^{\lambda s(X,y)}}{\sum_{f=1}^{k} e^{\lambda s(X,y_f)}}}{\frac{e^{\lambda s(X',y)}}{\sum_{f=1}^{k} e^{\lambda s(X',y_f)}}} \\ &= \frac{e^{\lambda s(X,y)}}{e^{\lambda s(X',y)}} \frac{\sum_{f=1}^{k} e^{\lambda s(X',y_f)}}{\sum_{f=1}^{k} e^{\lambda s(X,y_f)}} \\ &= e^{\lambda(s(X,y)-s(X',y))} \frac{\sum_{f=1}^{k} e^{\lambda s(X',y_f)}}{\sum_{f=1}^{k} e^{\lambda s(X,y_f)}} \\ &\leq e^{\lambda \kappa d_1(X,X')} \frac{\sum_{f=1}^{k} e^{\lambda(\kappa d_1(X,X') + s(X,y_f))}}{\sum_{f=1}^{k} e^{\lambda s(X,y_f)}} \\ &= e^{2\lambda\kappa d_1(X,X')} \\ &\leq e^{d_1(X,X')\epsilon} \end{aligned}$$

for all datasets $X, X'$. Similarly $\frac{P(M_{s,E(\lambda)}(X')=y)}{P(M_{s,E(\lambda)}(X)=y)} \leq e^{d_1(X,X')\epsilon}$. Hence $|\ln(P(M_{s,E(\lambda)}(X) = y)) - \ln(P(M_{s,E(\lambda)}(X') = y))| \leq d_1(X, X')\epsilon$ as desired. □

We compute how well $s(X, M_{s,E(\lambda)}(X))$ is concentrated around $s(X, y_X)$.

**Proposition 3.12.** For finite $Y$ and $\mathbb{N}^{|\mathcal{X}|} \times Y \xrightarrow{s} \mathbb{R}$, its exponential mechanism $M_{s,E(\lambda)}$ satisfies

$$P\left(s(X, y_X) - s(X, M_{s,E(\lambda)}(X)) \geq \frac{1}{\lambda}\left(\ln\left(\frac{|Y|}{|Y_X|}\right) + \delta\right)\right) \leq e^{-\delta}$$

for every dataset $X$.

*Proof.* Write $c = s(X, y_X) - \frac{1}{\lambda}\left(\ln\left(\frac{|Y|}{|Y_X|}\right) + \delta\right)$. By definition

$$\begin{aligned}
P(s(X, M_{s,E(\lambda)}(X)) \leq c) &= \sum_{y \in Y} P(M_{s,E(\lambda)}(X) = y, s(X, y) \leq c) \\
&= \sum_{y \in Y, s(X,y) \leq c} P(M_{s,E(\lambda)}(X) = y) \\
&= \sum_{y \in Y, s(X,y) \leq c} \frac{e^{\lambda s(X,y)}}{\sum_{f=1}^{k} e^{\lambda s(X, y_f)}} \\
&\leq \frac{|Y| e^{\lambda c}}{\sum_{f=1}^{k} e^{\lambda s(X, y_f)}} \\
&\leq \frac{|Y| e^{\lambda c}}{|Y_X| e^{\lambda s(X, y_X)}} \\
&= \frac{|Y|}{|Y_X|} e^{\lambda(c - s(X, y_X))} \\
&= \frac{|Y|}{|Y_X|} e^{\lambda(-\frac{1}{\lambda}(\ln\left(\frac{|Y|}{|Y_X|}\right) + \delta))} \\
&= e^{-\delta}
\end{aligned}$$

□

**Corollary 3.13.** For finite $Y$ and $(\mathbb{N}^{|\mathcal{X}|} \times Y, d_1) \xrightarrow{s} (\mathbb{R}, d_1)$ with modified Lipschitz constant $\kappa$, if $\lambda = \frac{\epsilon}{2\kappa}$ then its exponential mechanism $M_{s,E(\lambda)}$ is $\epsilon$-differentially private and satisfies

$$P\left(s(X, y_X) - s(X, M_{s,E(\lambda)}(X)) \geq \frac{2\kappa}{\epsilon}\left(\ln\left(\frac{|Y|}{|Y_X|}\right) + \delta\right)\right) \leq e^{-\delta}$$

for every dataset $X$. In particular,

$$P\left(s(X, y_X) - s(X, M_{s,E(\lambda)}(X)) \geq \frac{2\kappa}{\epsilon}\left(\ln(|Y|) + \delta\right)\right) \leq e^{-\delta}$$

for every dataset $X$.

*Proof.* The first statement follows immediately from proposition 3.11 and proposition 3.12. The second statement follows from the first statement and our initial assumption $|Y_X| \geq 1$. □

3.4. **Gaussian Mechanism.** We can use Gaussian random variable to create random mechanisms such that 1 and 2 hold.

**Definition 3.14.** For $\sigma > 0$ and $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} \mathbb{R}^k$, we define its Gaussian mechanism

$$\begin{aligned} \mathbb{N}^{|\mathcal{X}|} &\xrightarrow{M_{f,N(\sigma)}} RV((\Omega, \mathcal{F}, P), (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))) \\ X &\mapsto f(X) + (N(\sigma), \dots, N(\sigma)) \end{aligned}$$

where $(\Omega, \mathcal{F}) \xrightarrow{N(\sigma)} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is the Gaussian random variable with probability density function $f_{N(\sigma)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$.

**Proposition 3.15.** For $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}^k, d_2)$ with Lipschitz constant $\kappa$, if $\sigma \geq \sqrt{2\ln(\frac{5}{4\delta})}\frac{\kappa}{\epsilon}$ then its Gaussian mechanism $M_{f,N(\sigma)}$ is $(\epsilon, \delta)$-differentially private.

*Proof.* mimic [4, appendix A]. □

We compute how well $M_{f,N(\sigma)}(X)$ is concentrated around $f(X)$.

**Proposition 3.16.** For $\delta \in (0, 1]$ and $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} \mathbb{R}^k$, its Gaussian mechanism $M_{f,N(\sigma)}$ satisfies

$$P(||M_{f,N(\sigma)}(X) - f(X)||_\infty \geq \text{ something}) \leq \delta$$

for every dataset $X$.

*Proof.* todo □

**Corollary 3.17.** For $\delta \in (0, 1]$ and $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}^k, d_2)$ with Lipschitz constant $\kappa$, if $\sigma = \sqrt{2\ln(\frac{5}{4\delta})}\frac{\kappa}{\epsilon}$ then its Gaussian mechanism $M_{f,N(\sigma)}$ is $(\epsilon, \delta)$-differentially private and satisfies

$$P(||M_{f,N(\sigma)}(X) - f(X)||_\infty \geq \text{ something}) \leq \delta$$

for every dataset $X$.

*Proof.* This follows immediately from proposition 3.15 and proposition 3.16. □

See example 5.2 for an illustration of the Gaussian mechanism.

3.5. **Localized Gaussian mechanism.** Instead of adding the same noise to each $f(X), X \in \mathbb{N}^{|\mathcal{X}|}$ in Gaussian mechanism, one can add less noise specific to each $X$ to create an equally differentially private mechanism that is more accurate.

**Definition 3.18.** For $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{\sigma} (0, \infty)$ and $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} \mathbb{R}^k$, we define its Gaussian mechanism

$$\begin{aligned} \mathbb{N}^{|\mathcal{X}|} &\xrightarrow{M_{f,N(\sigma)}} RV((\Omega, \mathcal{F}, P), (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))) \\ X &\mapsto f(X) + (N(\sigma(X)), \dots, N(\sigma(X))) \end{aligned}$$

where $(\Omega, \mathcal{F}) \xrightarrow{N(\sigma(X))} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is the Gaussian random variable with probability density function $f_{N(\sigma(X))}(x) = \frac{1}{\sqrt{2\pi\sigma(X)^2}} e^{-\frac{x^2}{2\sigma(X)^2}}$.

**Proposition 3.19.** For $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{\sigma} (0,\infty)$ and $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}^k, d_2)$ with localized Lipschitz constant $\kappa(X)$, if $\sigma(X) \geq \sqrt{2\ln(\frac{5}{4\delta})}\frac{\kappa(X)}{\epsilon}$ then its Gaussian mechanism $M_{f,N(\sigma)}$ is $(\epsilon,\delta)$-differentially private.

*Proof.* todo: mimic [4, appendix A] while confirming the right definition of localized Lipschitz constant. □

We compute how well $M_{f,N(\sigma)}(X)$ is concentrated around $f(X)$.

**Proposition 3.20.** For $\delta \in (0,1]$, $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{\sigma} (0,\infty)$ and $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} \mathbb{R}^k$, its Gaussian mechanism $M_{f,N(\sigma)}$ satisfies

$$P(||M_{f,N(\sigma)}(X) - f(X)||_\infty \geq \text{ something}) \leq \delta$$

for every dataset $X$.

*Proof.* todo □

**Corollary 3.21.** For $\delta \in (0,1]$, $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{\sigma} (0,\infty)$ and $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}^k, d_2)$ with localized Lipschitz constant $\kappa(X)$, if $\sigma(X) = \sqrt{2\ln(\frac{5}{4\delta})}\frac{\kappa(X)}{\epsilon}$ then its Gaussian mechanism $M_{f,N(\sigma)}$ is $(\epsilon,\delta)$-differentially private and satisfies

$$P(||M_{f,N(\sigma)}(X) - f(X)||_\infty \geq \text{ something}) \leq \delta$$

for every dataset $X$.

*Proof.* This follows immediately from proposition 3.19 and proposition 3.20. □

todo: see if this approach of adding less noise specific to each $X$ applies to Laplace mechanism, exponential mechanism as well.

3.6. **Sample-and-aggregate mechanism.** todo: partition $X = \coprod_{f=1}^{k} X_f$ (sample step) and replace $f(X)$ with $f'(X) = g(f(X_1),\dots,f(X_k))$ (aggregate step) that has smaller $\kappa, \kappa(X)$ (so that less noise is needed) and is close to $f(X)$ (so that accuracy stays the same).

We conclude this section by stressing that the theoretical bounds for privacy and accuracy in the above propositions are upper bounds, so the privacy and accuracy in practice are usually better.

## 4. In Data Science

4.1. **Response.** For the response function $\{0,1\}^k \xrightarrow{\text{id}} \{0,1\}^k$ as in example 2.10, its randomized response mechanism

$$\begin{aligned} \{0,1\}^k &\xrightarrow{M_{\text{id},B(p),B(q)}} RV((\Omega,\mathcal{F},P),(\{0,1\}^k,\mathcal{G})) \\ (\dots,b_f,\dots) &\mapsto M_{\text{id},B(p),B(q)}(\dots,b_f,\dots) \end{aligned}$$

is $\max\left\{\frac{1-p}{q}, \frac{p}{1-q}, \frac{q}{1-p}, \frac{1-q}{p}\right\}$-differentially private and satisfies

$$P(||M_{\mathrm{id},B(p),B(q)}(\dots,b_f,\dots)-(\dots,b_f,\dots)||_1 = j) \leq C(k,j)\max\{p,q\}^j \max\{1-p,1-q\}^{k-j}$$

for every response $(\dots,b_f,\dots)$ by corollary 3.4.

**Example 4.1.** Consider the response function $\{0,1\} \xrightarrow{\mathrm{id}} \{0,1\}$. It has Lipschitz constant $\kappa = 1$.

a. for $\epsilon = 0, p = \frac{1}{2}, q = \frac{1}{2}$, its randomized response mechanism $M_{\mathrm{id},B(p),B(q)}$ is 0-differentially private and satisfies
$$P(|M_{\mathrm{id},B(p),B(q)}(b) - b| = 1) = \frac{1}{2}$$
for every bit $b$. This is the case of perfect privacy and zero accuracy in example 2.23.

b. for $\epsilon = \ln(3), p = \frac{1}{4}, q = \frac{1}{4}$, its randomized response mechanism $M_{\mathrm{id},B(1/4),B(1/4)}$ is $\ln(3)$-differentially private and satisfies
$$P(|M_{\mathrm{id},B(p),B(q)}(b) - b| = 1) = \frac{1}{4}$$
for every bit $b$. This is [4, subsection 3.2].

## 5. In Machine Learning

We borrow from the review in [7, section 2]. A common application of differential privacy in machine learning is to produce differentially private approximation to the machine learning model at hand. To do this, one can

- add noise to the trained model. This is similar to $f(X) + (L(\beta),\dots,L(\beta))$ in subsection 3.2 and is sometimes called output perturbation.
- add noise to the objective function. This is similar to $g(f(X)+(L(\beta),\dots,L(\beta))$ in subsection 3.2 plus subsection 2.2 and is sometimes called objective perturbation.
- add noise to each iteration in the algorithm if it is iterative.
- sample from the dataset to produce different models before adding noise to their aggregate. This is similar to what is in subsection 3.6.
- others

Afterward one can measure the performance of his differentially private mechanism by

- choosing what the true model is
- choosing a distance between models
- measuring the distance between the differentially private approximation and the true model
- analyzing convergence of this distance with respect to the size of the dataset

We go through how all this is done for some machine learning algorithms to get a better sense.

### 5.1. Baysian conjugate models.

We can produce differentially private approximations to Bayesian conjugate models by either adding noise to each closed-form update or by adding noise to the trained models (output perturbation) and provide theoretical bounds and experimental numbers.

5.2. **Counting.** For the counting function $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} \mathbb{R}$ in example 2.12 with Lipschitz constant $\kappa = 1$, its Laplace mechanism

$$\begin{aligned} \mathbb{N}^{|\mathcal{X}|} &\xrightarrow{M_{f,L(1/\epsilon)}} RV((\Omega, \mathcal{F}, P), (\mathbb{R}, \mathcal{B}(\mathbb{R}))) \\ X &\mapsto f(X) + L(1/\epsilon) \end{aligned}$$

is $\epsilon$-differentially private and satisfies

$$P\left(|M_{f,L(\beta)}(X) - f(X)| \geq \frac{1}{\epsilon} \ln\left(\frac{1}{\delta}\right)\right) \leq \delta$$

for every dataset $X$ by corollary 3.8.

**Example 5.1.** Consider the binary function $\mathcal{X} \xrightarrow{b} \{0, 1\}, x \mapsto 1$ and its counting function $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}, d_1), X \mapsto \sum_{x \in X} b(x)$ that counts the number of elements in each dataset. It has Lipschitz constant $\kappa = 1$.

a. for $\epsilon = 1, \beta = 1, \delta = 0.1$, its Laplace mechanism $M_{f,L(1)}$ is 1-differentially private and satisfies
$$P(|M_{f,L(\beta)}(X) - f(X)| \geq \ln 10) \leq 0.1$$
for every dataset $X$.

b. for $\epsilon = 1, \beta = 1, \delta = \frac{1}{e}$, its Laplace mechanism $M_{f,L(1)}$ is 1-differentially private and satisfies
$$P(|M_{f,L(\beta)}(X) - f(X)| \geq 1) \leq \frac{1}{e}$$
for every dataset $X$.

5.3. **Decision Tree.** see [6] for how to produce differentially private approximation to decision tree. It does so by randomly partitioning the sample space into $P_{i1}, \ldots, P_{in_i}$ for each tree $T_i, 1 \leq i \leq m$, counting the number $c_{ijk}$ of samples of label $k$ in each partition $P_{ij}$, and predicting based on $P(Y = k \,|\, X = x) = \frac{\sum_{i=1, x \in P_{ij_i}}^{m} c_{ij_ik}}{\sum_k \sum_{i=1, x \in P_{ij_i}}^{N} c_{ij_ik}}$.

5.4. **Histogram.** For the histogram function $\mathbb{N}^{|\mathcal{X}|} \xrightarrow{f} \mathbb{R}^k$ in example 2.13 with Lipschitz constant $\kappa = 1$, its Laplace mechanism

$$\begin{aligned} \mathbb{N}^{|\mathcal{X}|} &\xrightarrow{M_{f,L(1/\epsilon)}} RV((\Omega, \mathcal{F}, P), (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))) \\ X &\mapsto f(X) + (L(1/\epsilon), \ldots, L(1/\epsilon)) \end{aligned}$$

is $\epsilon$-differentially private and satisfies

$$P\left(||M_{f,L(\beta)}(X) - f(X)||_\infty \geq \frac{1}{\epsilon} \ln\left(\frac{k}{\delta}\right)\right) \leq \delta$$

for every dataset $X$ by corollary 3.8.

**Example 5.2.** todo: Consider the binning function $\mathcal{X} \xrightarrow{b} \{0,1\}^k, x_i \mapsto (0,\dots 0,1,0,\dots,0)$ and its histogram function $(\mathbb{N}^{|\mathcal{X}|}, d_1) \xrightarrow{f} (\mathbb{R}^k, d_1), X \mapsto \sum_{i=1}^{n} b(x_i)$. It has Lipschitz constant $\kappa = 1$.
a. for $\epsilon = 1, \sigma = 1, \delta = 0.1$, its Gaussian mechanism $M_{f,G(1)}$ is...
b. for $\epsilon = 1, \sigma = 1, \delta = \frac{1}{\epsilon}$, its Gaussian mechanism $M_{f,G(\frac{1}{\epsilon})}$ is...

5.5. **K-means Clustering.** see [9] for how to produce differentially private approximation to $K$-means clustering by using the smooth sensitivity mechanism and the sample-and-aggregate mechanism.

5.6. **Linear Regression.** see [11] for how to produce differentially private approximation to linear regressor. It does so by approximating the objection function with its low-order Taylor expansion and adding Laplace noise to the Taylor expansion coefficients.

5.7. **Logistic Regression.** see [1] for how to produce differentially private approximation to logistic regressor. It does so by adding noise to the output or by adding noise to the objective function.

5.8. **Naive Bayes.** see [10] for how to produce differentially private approximation to naive Bayes classifier. It does so by adding Laplace noise to $P(X_i = x_i \,|\, Y = y_k)$ in $P(Y = k \,|\, X_1 = x_1, \dots, X_m = x_m) \propto P(Y = k) \prod_{i=1}^{m} P(X_i = x_i \,|\, Y = k)$.

5.9. **Neural Networks.** see [8] for how to produce differentially private approximation to neural network. It does so by sampling from the batch or mini-batch, clipping the gradients, adding noise to the clipped gradients, averaging the noised clipped gradients before updating model parameters.

5.10. **SVM.** see [2] for how to produce differentially private approximations to classifiers learned via empirical risk minimization such as SVM. It does so by adding noise to the output or by adding noise to the objective function.

## 6. Others

- compliance to GDPR
- market demand in US
- produce each differentially private approximation to each model in its own package or produce them all in a `differential_privacy` package

## 7. Data trail

- local privacy instead of central privacy: data is privatized before being sent
- at event level: each event is privatized
- with limit: fixed number of events per time unit
- anonymity: no name, no IP identifier, etc. : even pairing 2 batches is impossible
- protection: encrypted channel and restricted access
- transparency: opt in or opt out

## References


[1] Chaudhuri K. and Monteleoni C., *Privacy-preserving logistic regression*, Advances in Neural Information Processing Systems (2008), 289–296.

[2] Chaudhuri K., Monteleoni C., and Sarwate A. D., *Differentially private empirical risk minimization*, Journal of Machine Learning Research (2011), 1069–1109.

[3] I. Dinur and K. Nissim, *Revealing information while preserving privacy*, Proceedings of the 22th ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 2003, pp. 202–210.

[4] Dwork C. and Roth A., *The Algorithmic Foundations of Differential Privacy*, Foundations and Trends in Theoretical Computer Science **9** (2014), no. 3–4, 211–407.

[5] Dwork C., McSherry F., Nissim K., and Smith A., *Calibrating Noise to Sensitivity in Private Data Analysis*, Lecture Notes in Computer Science, 2006, pp. 265–284.

[6] Jagannathan G., Pillaipakkamnatt K., and Wright R. N., *A practical differentially private random decision tree classifier*, International Conference on Data Mining Workshops (2009), 114-121.

[7] Ji Z., Lipton Z. C., and Elkan C., *Differential privacy and machine learning: A survey and review* (2014), available at `https://arxiv.org/abs/1412.7584`.

[8] McMahan B., Andrew G., and Erlingsson U., *A general approach to adding differential privacy to iterative training Procedures* (2018), available at `https://arxiv.org/abs/1812.06210`.

[9] Nissim K. and Smith A., *Smooth sensitivity and sampling in private data analysis*, ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (2007), 75–84.

[10] Vaidya J., Shafiq B., Basu A., and Hong Y., *Differentially Private Naive Bayes Classification*, Web Intelligence **1** (2013), 571–576.

[11] Zhang J., Zhang Z., Xiao X., Yang Y., and Winslett M., *Functional mechanism: Regression analysis under differential privacy*, International Conference on Very Large Data Bases (2012), 1364-1375.