Note on Differential Privacy

Anderson Lee

March 21, 2025

Abstract

This note is taken when studying [DR14].

Contents

Chapter 1

Basic Terms

1.1 Formalizing Differential Privacy

Definition 1.1.1 (Probability Simplex). Given a discrete set B, the *probability simplex* over B, denoted $\Delta(B)$, is defined to be:

$$\Delta(B) = \left\{ x \in \mathbb{R}^{|B|} : x_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{|B|} x_i = 1 \right\}$$

Definition 1.1.2 (Randomized Algorithm). A randomized algorithm $\mathcal{M}$ with domain A and discrete range B is associated with a mapping $M : A \to \Delta(B)$. On input $a \in A$, the algorithm $\mathcal{M}$ outputs $\mathcal{M}(a) = b$ with probability $(M(a))_b$ for each $b \in B$. The probability space is over the coin flips of the algorithm $\mathcal{M}$.

A database x is a collection of records from a universe $\mathcal{X}$. The representation is usually a histogram: $x \in \mathbb{N}^{|\mathcal{X}|}$ where each entry x_i represents the number of elements in the database x of *type* $i \in \mathcal{X}$. For instance, let $\mathcal{X} = \{A, B, C, D\}$ and $x = [0, 2, 2, 0] \in \mathbb{N}^{|4|}$. That means database x contains 2 records of B and 2 records of C.

Definition 1.1.3 (Distance Bewteen Databases). The distance between two databases x, y is by the ℓ_1 norm defined as:

$$\|x - y\|_1 = \sum_{i=1}^{|\mathcal{X}|} |x_i - y_i|$$

Definition 1.1.4 (Differential Privacy). A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ε, δ)-differentially private if for all $\mathcal{S} \subset \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|X|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr\left[\mathcal{M}(x) \in \mathcal{S}\right] \leq \exp(\varepsilon) \Pr\left[\mathcal{M}(y) \in \mathcal{S}\right] + \delta$$

where the probability space is over the mechanism $\mathcal{M}$. If $\delta = 0$, we say $\mathcal{M}$ is ε-differentially private.

Typically, we are interested in a δ that's less than the inverse of any polynomial in the size of the database x.

Definition 1.1.5 (Privacy Loss). Given an output $\xi \sim \mathcal{M}$, the privacy loss is defined as:

$$\mathcal{L}^{(\xi)}_{\mathcal{M}(x) \| \mathcal{M}(y)} = \ln\left(\frac{\Pr\left[\mathcal{M}(x) = \xi\right]}{\Pr\left[\mathcal{M}(y) = \xi\right]}\right)$$

1.2 Immunity to Post-processing

Proposition 1.2.1 (Post-Processing). Let $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}$ be a randomized algorithm that is $(\varepsilon,\, \delta)$-differentially private. Let $f : R \to R'$ be an arbitrary randomized mapping. Then $f \circ \mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to R'$ is $(\varepsilon,\, \delta)$-differentially private.

Proof. We prove the proposition for a deterministic function $f : R \to R'$. The result then follows because **any randomized mapping can be decomposed into a convex combination of deterministic functions**, **and a convex combination of differentially private mechanism is also differentially private**.

Fix any pair of neighboring databases x, y with $\|x - y\|_1 \leq 1$, and fix any event $S \subset R'$. Let $T = \{r \in R : f(r) \in S\}$. We then have

$$\begin{aligned}
\Pr\left[f(\mathcal{M}(x)) \in S\right] &= \Pr\left[\mathcal{M}(x) \in T\right] \\
&\leq \exp(\varepsilon) \Pr\left[\mathcal{M}(y) \in T\right] + \delta \\
&= \exp(\varepsilon) \Pr\left[f(\mathcal{M}(y)) \in S\right] + \delta
\end{aligned}$$

which is what we wanted. ■

Remark. Differential privacy is immune to post-processing.

1.3 Economic View of DP's Promise

A differentially private mechanism $\mathcal{M}$ does not promise that an individual faces harm after the result is released. Instead, it promises that the probability of facing the harm is not significantly more than **not participating the database**. This provides an **incentive** for an individual to include their data in the database.

Consider an individual i who has arbitrary preference over the set of all possible future events, which we denote by $\mathcal{A}$. Let $u_i : \mathcal{A} \to \mathbb{R}^+$ be the utility function, where i experiences utility $u_i(a)$ when $a \in \mathcal{A}$ happens. Suppose x is a dataset containing private individuals' data, and y is a dataset identical to x except that it does not contain individual i's data. Suppose $\mathcal{M}$ is an ε-differentially privatealgorithm. Let $f : \mathrm{Range}(\mathcal{M}) \to \Delta(\mathcal{A})$ be the arbitrary function that determines the distribution over future events $\mathcal{A}$, conditioned on the output of $\mathcal{M}$. By the definition of differential privacy and its immunity to post-processing, we have

$$\begin{aligned}
\mathbb{E}_{a \sim f(\mathcal{M}(x))}\left[u_i(a)\right] &= \sum_{a \in \mathcal{A}} u_i(a) \cdot \Pr_{f(\mathcal{M}(x))}[a] \\
&\leq \sum_{a \in \mathcal{A}} u_i(a) \cdot \exp(\varepsilon) \cdot \Pr_{f(\mathcal{M}(y))}[a] \\
&= \exp(\varepsilon) \cdot \mathbb{E}_{a \sim f((M)(y))}[u_i(a)]
\end{aligned}$$

This promise that the expected utility is not harmed by more than a factor of $\exp(\varepsilon)$ regardless of participating the database. Additionally, this promise holds **independently** of the individual's utility function u_i, and holds **simultaneously** for multiple individuals who have completely different utility functions.

Chapter 2

Known Bugs

Lecture 2: Second Lecture

2.1 Introduction

9 Sep. 08:00

Nothing is bugs-free. There are some known bugs which I don't have incentive to solve, or it is hard to solve whatsoever. Let me list some of them.

2.1.1 Footnote Environment

It's easy to let you fall into a situation that you want to keep using `footnote` to add a bunch of unrelated stuffs. However, with our environment there is a known strange behavior, which is following.

Example. Footnote![a]

Remark. Oops! footnote somehow shows up earlier than expect![a]

[a] This is a footnote!

[a] This is another footnote!

Bugs caught![b]

[b] The final footnote which is ok!

As we saw, the footnote in the `Example` environment should show at the bottom of its own box, but it's caught by `Remark` which causes the unwanted behavior. Unfortunately, I haven't found a nice way to solve this. A potential way to solve this is by using `footnotemark` with `footnotetext` placing at the bottom of the environment, but this is tedious and needs lots of manual tweaking.

Furthermore, not sure whether you notice it or not, but the color box of `Remark` is not quite right! It extends to the right, another trick bug...

2.1.2 Mdframe Environment

Though `mdframe` package is nice and is the key theme throughout this template, but it has some kind of weird behavior. Let's see the demo.

Proof of ??. We need to prove the followings.

Claim. $E = mc^2$.

Proof. Nonsense.
Nonsense,
Nonsense,
Nonsense,
Nonsense,
Nonsense. ⊛

■

I expect it should break much earlier, and this seems to be an algorithmic issue of `mdframe`. One potential solution is to use `tcolorbox` instead, but I haven't completely figure it out, hence I can't really say anything right now.

Chapter 3

Test

Lecture 2: Second Lecture

3.1 Introduction

9 Sep. 08:00

Nothing is bugs-free. There are some known bugs which I don't have incentive to solve, or it is hard to solve whatsoever. Let me list some of them.

3.1.1 Footnote Environment

It's easy to let you fall into a situation that you want to keep using `footnote` to add a bunch of unrelated stuffs. However, with our environment there is a known strange behavior, which is following.

Example. Footnote![a]

Remark. Oops! footnote somehow shows up earlier than expect![a]

[a] This is a footnote!
[a] This is another footnote!

Bugs caught![b]

[b] The final footnote which is ok!

As we saw, the footnote in the `Example` environment should show at the bottom of its own box, but it's caught by `Remark` which causes the unwanted behavior. Unfortunately, I haven't found a nice way to solve this. A potential way to solve this is by using `footnotemark` with `footnotetext` placing at the bottom of the environment, but this is tedious and needs lots of manual tweaking.

Furthermore, not sure whether you notice it or not, but the color box of `Remark` is not quite right! It extends to the right, another trick bug...

3.1.2 Mdframe Environment

Though `mdframe` package is nice and is the key theme throughout this template, but it has some kind of weird behavior. Let's see the demo.

Proof of ??. We need to prove the followings.

Claim. $E = mc^2$.

Proof. Nonsense.
Nonsense,
Nonsense,
Nonsense,
Nonsense,
Nonsense. $\circledast$

$\blacksquare$

I expect it should break much earlier, and this seems to be an algorithmic issue of `mdframe`. One potential solution is to use `tcolorbox` instead, but I haven't completely figure it out, hence I can't really say anything right now.

Appendix

Appendix A

Additional Proofs

A.1 Proof of ??

We can now prove **??**.

Proof of ??. See here. ■

Bibliography

[DR14] Cynthia Dwork and Aaron Roth. “The Algorithmic Foundations of Differential Privacy”. In: *Foundations and Trends® in Theoretical Computer Science* 9.3–4 (2014), pp. 211–407. ISSN: 1551-305X. DOI: 10.1561/0400000042. URL: http://dx.doi.org/10.1561/0400000042.