ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

PRIVACY VISUALISATION THROUGH HYPOTHESIS TESTING

MATHEMATICS OF INFORMATION LABORATORY

Master's thesis - Salim Najib, supervised by Pr. Yanina Shkel and Cemre Çadir.

Spring 2025

ABSTRACT

Differential privacy has stood the test of time as a measure of information leakage, in a context where the privacy of individuals has become an increasingly critical stake. To help non-privacy experts get familiar with some of its recent generalisations and theoretical results on composition, we devise a visualisation tool for differential privacy, leveraging its binary hypothesis testing interpretation. It aims to show multiple kinds of privacy regions, to showcase differences between composition theorems, and to aid in gaining an intuition on privacy-utility trade-offs for some common mechanisms. The said tool can be found here [link]. In the process, we also investigate new research questions related to the composition of mechanisms for which multiple differential privacy constraints hold.

Contents

CHAPTER 1

INTRODUCTION

Preserving the privacy of individuals while presenting meaningful information is an extremely challenging yet important task, to all kinds of experts, domain-specific researchers and industry practitioners alike. Measuring the privacy versus utility trade-off alone remains an ongoing and very active topic of research. **Differential privacy** (DP) has emerged as one go-to mathematical measure of privacy.

Many variations of DP and other information-theoretic measures have been proposed, and much exciting research has been done on DP. Nonetheless, the gap from the formulas to their interpretations is often too large, even for statistically informed experts, scientists and engineers, to have a tangible grasp on their meanings and implications.

Consequently, in this project, we propose to tackle this problem via a **visualisation tool**, leveraging the **binary hypothesis testing interpretation** of differential privacy. Indeed, hypothesis testing is a ubiquitous framework in all fields related to statistics: biometrics, machine learning, data analysis, etc. Its universality will be leveraged to better understand well-established and recently proposed measures as well as their degradation under query composition.

This work is inspired by Panavas et al., *A Visualization Tool to Help Technical Practitioners of Differential Privacy* [1]. The novelty is essentially three-fold: the focus on the binary hypothesis testing interpretation of DP, the incorporation of visualisations of complex composition theorems for DP and some of its variations, and the plotting of specific mechanisms' privacy regions along with the comparison to their theoretical DP or DP variant guarantees.

As a Master's thesis, the main goals of this project are the following:

- Gaining familiarity with differential privacy, its recent variations and the state of the art of research on DP ($\sim$ chapter 2).
- Implementing a visualisation tool utilising the binary hypothesis testing perspective on differential privacy ($\sim$ chapter 3).
- Thinking of new lines of research in DP and beyond ($\sim$ chapter 4).

All the code relevant to this project can be found on `https://github.com/Dicedead/privacyVis`.

CHAPTER 2

THEORETICAL BACKGROUND

This chapter motivates and introduces the basics of differential privacy, along with some of its relatively recent variations and theoretical results on the composition of differentially private mechanisms.

2.1 INTRODUCING DIFFERENTIAL PRIVACY

We will follow the same introduction to differential privacy as Pr. Kamath presents in his introductory lecture on the topic [2].

2.1.1 MOTIVATION: ARE YOU A CHEATER?

Assume you suspect that a fraction p of the n students in your class are cheating. How can you estimate p? The natural idea is to ask your n students whether they are cheating or not, theoretically obtaining n responses

$$\{X_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} \text{Ber}\,(p)\,,$$

where $X_i = 1$ if the i^{th} student is cheating, and $X_i = 0$ if they are not. Then, we can estimate p as

$$\hat{p}_0 = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

However, no student would want to reveal the *sensitive* information that they are cheating, thus this approach cannot work. For students to answer truthfully, they need *plausible deniability*, i.e a way to answer truthfully without revealing that they are cheating, if they are.

A simple initial approach is the following:

1. Ask each student if they are cheating or not, and to keep the answer $x \in \{0, 1\}$ for themselves,
2. then flip a coin.

3a. If the coin lands on heads, report x, their initial response.

3b. If the coin lands on tails, report $1 - x$, the opposite of their initial response.

This coincides with measuring

$$Y_i = \begin{cases} X_i & \text{w.p } \frac{1}{2}, \\ 1 - X_i & \text{w.p } \frac{1}{2} \end{cases}$$

then estimating p from

$$\hat{p}_{\frac{1}{2}} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Unfortunately, $\{Y_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} \text{Ber}\left(\frac{1}{2}\right)$ thus $n\hat{p}_{\frac{1}{2}} \sim \text{Bin}\left(n, \frac{1}{2}\right)$ is an ancillary statistic w.r.t p. In simpler terms, the distribution of $n\hat{p}_{\frac{1}{2}}$ does not depend on p. Thus, while this approach yields no sensitive information, it also provides no sensible estimate of p.

2.1.2 BINARY RANDOMIZED RESPONSE

One viable approach is to interpolate between the two previous naive solutions. Given some *privacy budget* parametrized by $\gamma \in \left[0, \frac{1}{2}\right]$, one observes the following randomized functions Y_i from X_i,

$$Y_i = \begin{cases} X_i & \text{w.p } \frac{1}{2} + \gamma, \\ 1 - X_i & \text{w.p } \frac{1}{2} - \gamma. \end{cases}$$

For $\gamma \in \left\{0, \frac{1}{2}\right\}$, this corresponds to the two naive strategies explained above. For $\gamma \in \left]0, \frac{1}{2}\right[$, we get an estimator for p

$$\hat{p}_\gamma = \frac{1}{2\gamma n}\sum_{i=1}^{n}\left(Y_i - \frac{1}{2} + \gamma\right)$$

s.t $\mathbb{E}\left(\hat{p}_\gamma\right) = p$ and, by the Chebyshev inequality,

$$\mathbb{P}\left(|\hat{p}_\gamma - p| \geq \varepsilon\right) \leq \frac{1}{16\gamma^2\varepsilon^2 n}.$$

The RHS is a measure for the accuracy of $\hat{p}_\gamma$.

In the previous exposition, we have assumed γ to be some meaningful parametrization of privacy in this context. Can we quantify how privacy-preserving this system is more operationally? Differential privacy is one possible answer.

2.1.3 DEFINITION OF DIFFERENTIAL PRIVACY

First we need to define the notion of **neighbouring datasets**.

DEFINITION 1: NEIGHBOURING DATASETS
Let $\mathcal{X}$ be a set and $X, X' \in \mathcal{X}^n$. X and X' are **neighbouring datasets** if they differ in exactly one entry.

This definition of differential privacy is taken from [3].

DEFINITION 2: DIFFERENTIAL PRIVACY, (ε, δ)-DP
Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a randomized function/mechanism/algorithm, $\varepsilon \geq 0$ and $\delta \in [0, 1]$.

M is (ε, δ) **−differentially private** or (ε, δ) **−DP** if $\forall X, X' \in \mathcal{X}^n$ which are neighbouring and all $T \subseteq \mathcal{Y}$,

$$\mathbb{P}\left(M(X) \in T\right) \leq e^{\varepsilon}\mathbb{P}\left(M(X') \in T\right) + \delta.$$

If $\delta = 0$, we also say that M is $\varepsilon-$(pure) differentially private, or $\varepsilon-$(pure) DP.
If $\delta > 0$, we sometimes speak of **approximate** differential privacy.

Intuitively, DP is a worst case measure of how likely it is to guess correctly that the underlying dataset producing the observable $Y \in \{M(X), M(X')\}$ is X or X'. Since X and X' differ in only one position, this guess provides significant (sensitive) insight into that position. Later on, we will justify this intuition through a hypothesis testing perspective, in section 2.2.

As an example, we estimate the differential privacy of the binary randomized response mechanism, presented in the previous section, with $M(X) = M(X_1^n) = Y_1^n$. W.l.o.g assume $X, X' \in \{0, 1\}^n$ differ only in position 1. Let $y \in \{0, 1\}^n$ and $\gamma \in \left]0, \frac{1}{4}\right[$. Then since the Y_i are independent,

$$\begin{aligned}
\frac{\mathbb{P}\left(M(X) = y\right)}{\mathbb{P}\left(M(X') = y\right)} &= \frac{\mathbb{P}\left(Y_1 = y_1, \dots Y_n = y_n\right)}{\mathbb{P}\left(Y_1' = y_1, \dots, Y_n' = y_n\right)} \\
&= \frac{\mathbb{P}\left(Y_1 = y_1\right)}{\mathbb{P}\left(Y_1' = y_1\right)} \\
&\leq \frac{\frac{1}{2} + \gamma}{\frac{1}{2} - \gamma} \\
&= \frac{1 + 2\gamma}{1 - 2\gamma} \\
&\overset{\gamma < \frac{1}{4}}{<} 2(1 + 2\gamma) \\
&\leq 2e^{2\gamma} \\
&= e^{2\gamma + \ln(2)}.
\end{aligned}$$

Thus for $\gamma \in \left]0, \frac{1}{4}\right[$, binary $\gamma-$randomized response is $(2\gamma + \ln(2))$ $-$pure DP.

2.1.4 STRUCTURAL PROPERTIES OF DIFFERENTIAL PRIVACY

We have yet to prove that the final estimate of the proportion of cheaters $\hat{p}_\gamma$ is privacy-preserving. Fortunately, DP is conserved under function composition.

PROPOSITION 3: CONSERVATION UNDER POST-PROCESSING
Let $M : \mathcal{X}^n \to \mathcal{Y}$ be $(\varepsilon, \delta)-$DP, and let $F : \mathcal{Y} \to \mathcal{Z}$ be a possibly randomized mapping.
Then $F \circ M$ is $(\varepsilon, \delta)-$DP.

The proof of this result is in appendix B.1.1. For completeness, we also include a simple result on datasets differing by k positions. Its proof can also be found in B.1.1.

COROLLARY 4: GROUP PRIVACY
Let $M : \mathcal{X}^n \to \mathcal{Y}$ be an $(\varepsilon, \delta)-$DP mechanism, and $X, X' \in \mathcal{X}^n$ differing in exactly k positions. Then for all $T \subseteq \mathcal{Y}$,

$$\mathbb{P}\left(M(X) \in T\right) \leq e^{k\varepsilon}\mathbb{P}\left(M(X') \in T\right) + ke^{k-1}\delta.$$

2.2 HYPOTHESIS TESTING PERSPECTIVE ON DIFFERENTIAL PRIVACY

In the following, we will need a very important operational perspective on differential privacy, mentioned in [4] and further explained and used in [5]. Can we understand, in statistical terms, what problem differential privacy represents? The answer is yes and it is the following.

2.2.1 RELATIONSHIP BETWEEN DIFFERENTIAL PRIVACY AND HYPOTHESIS TESTING

Let $X_0, X_1 \in \mathcal{X}^n$ be two neighbouring datasets and $M : \mathcal{X}^n \to \mathcal{Y}$ a $(\varepsilon, \delta)-$DP mechanism. The adversary observes $Y \in \{M(X_0), M(X_1)\}$ without the knowledge of the input to M. Hence the adversary has two hypotheses:

$$H_0 : Y = M(X_0), \text{ i.e the underlying dataset is } X_0, \text{ or}$$
$$H_1 : Y = M(X_1), \text{ i.e the underlying dataset is } X_1.$$

Then the adversary's task is to design a test region $T \subseteq \mathcal{Y}$ such that she rejects H_0 if $Y \in T$, and she cannot reject H_0 if $Y \in \mathcal{Y} \setminus T = T^C$. T should be chosen as to trade-off the following two types of errors:

$$\text{Type I error (false positive): FP}\left(X_0, X_1, M, T\right) = \mathbb{P}\left(Y \in T | H_0\right) = \mathbb{P}\left(M(X_0) \in T\right),$$
$$\text{Type II error (false negative): FN}\left(X_0, X_1, M, T\right) = \mathbb{P}\left(Y \in T^C | H_1\right) = \mathbb{P}\left(M(X_1) \in T^C\right).$$

The natural question becomes: seen as M is $(\varepsilon, \delta)-$DP, what can we infer on the false positive and the false negative probabilities, and conversely?

THEOREM 5: RELATING DIFFERENTIAL PRIVACY TO HYPOTHESIS TESTING
Let $\varepsilon \geq 0, \delta \in [0, 1]$ and $M : \mathcal{X}^n \to \mathcal{Y}$ be a mechanism. Then

$$M \text{ is } (\varepsilon, \delta) - \text{DP}$$
$$\iff$$
$\forall$ neighbouring $X_0, X_1 \in \mathcal{X}^n \; \forall T \subseteq \mathcal{Y}$,

$$\text{FP}\left(X_0, X_1, M, T\right) + e^{\varepsilon}\text{FN}\left(X_0, X_1, M, T\right) \geq 1 - \delta \text{ and}$$
$$e^{\varepsilon}\text{FP}\left(X_0, X_1, M, T\right) + \text{FN}\left(X_0, X_1, M, T\right) \geq 1 - \delta.$$

PROOF: Unpacking the notation, the equivalence becomes

$$M \text{ is } (\varepsilon, \delta) - \text{DP}$$
$$\iff$$

$\forall$ neighbouring $X_0, X_1 \in \mathcal{X}^n \; \forall T \subseteq \mathcal{Y}$,

$$\mathbb{P}\left(M(X_0) \in T\right) + e^{\varepsilon}\mathbb{P}\left(M(X_1) \in T^C\right) \geq 1 - \delta \text{ and}$$
$$e^{\varepsilon}\mathbb{P}\left(M(X_0) \in T\right) + \mathbb{P}\left(M(X_1) \in T^C\right) \geq 1 - \delta.$$

Let $X_0, X_1 \in \mathcal{X}^n$ be neighbouring datasets and $T \subseteq Y$.

$\Longrightarrow$: First observe that, because M is $(\varepsilon, \delta)-$DP,

$$\mathbb{P}\left(M(X_1) \in T^C\right) \geq \frac{\mathbb{P}\left(M(X_0) \in T^C\right) - \delta}{e^{\varepsilon}}.$$

Thus

$$\begin{aligned}\mathbb{P}\left(M(X_0) \in T\right) + e^{\varepsilon}\mathbb{P}\left(M(X_1) \in T^C\right) &\geq \mathbb{P}\left(M(X_0) \in T\right) + e^{\varepsilon}\frac{\mathbb{P}\left(M(X_0) \in T^C\right) - \delta}{e^{\varepsilon}} \\ &= \mathbb{P}\left(M(X_0) \in T\right) + \mathbb{P}\left(M(X_0) \in T^C\right) - \delta \\ &= 1 - \delta,\end{aligned}$$

and we obtain the second proposition that we have to prove by swapping the roles of X_0 and X_1 and replacing T by T^C.

$\Longleftarrow$: The second line in the equivalence can be written as follows:

$$e^{\varepsilon}\mathbb{P}\left(M(X_0) \in T\right) + \delta \geq 1 - \mathbb{P}\left(M(X_1) \in T^C\right) = \mathbb{P}\left(M(X_1) \in T\right).$$

Since X_0, X_1 are arbitrary neighbouring datasets, this proves that M is $(\varepsilon, \delta)-$DP. □

2.2.2 PRIVACY REGIONS

The previous theorem suggests two dual visualizations of differential privacy. The first one is the notion of a **privacy region for** $(\varepsilon, \delta)-$**DP**, which captures, for a given DP budget (ε, δ), the achievable false positive and false negative rates for the hypothesis test explained above.

DEFINITION 6: PRIVACY REGION FOR $(\varepsilon, \delta)-$DP
Let $\varepsilon \geq 0$ and $\delta \in [0, 1]$. The **privacy region for** $(\varepsilon, \delta)-$**DP** is

$$\mathcal{R}\left(\varepsilon, \delta\right) = \left\{ \left(\text{FN}, \text{FP}\right) \in [0,1]^2 \;\middle|\; \begin{cases} \text{FP} + e^{\varepsilon}\text{FN} \geq 1 - \delta, \\ e^{\varepsilon}\text{FP} + \text{FN} \geq 1 - \delta. \end{cases} \right\}.$$

The second one is the notion of a **privacy region of a mechanism** M, firstly with respect to neighbouring datasets X_0, X_1. This encompasses the false positive and false negative rates achievable by a mechanism for two given neighbouring datasets.

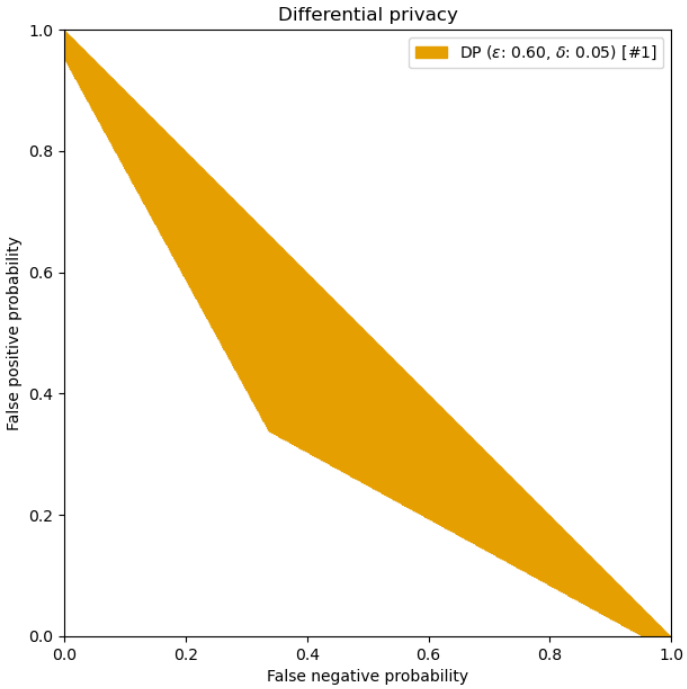

FIGURE 2.1
(0.6, 0.05)-DP region

DEFINITION 7: PRIVACY REGION FOR A MECHANISM (AND TWO NEIGHBOURING DATASETS)
Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a mechanism and let X_0, X_1 be two neighbouring datasets in $\mathcal{X}^n$. Then the **privacy region for** (M, X_0, X_1) is

$$\mathcal{R}(M, X_0, X_1) = \text{conv}\left\{(\text{FN}(X_0, X_1, M, T), \text{FP}(X_0, X_1, M, T)) \mid T \subseteq \mathcal{Y}\right\}.$$

We can then naturally define the **privacy region of a mechanism**:

$$\mathcal{R}(M) = \bigcup_{\substack{X_0, X_1 \in \mathcal{X}^n \\ X_0, X_1 \text{ neighbouring}}} \mathcal{R}(M, X_0, X_1).$$

We immediately get the following visual corollary.

COROLLARY 8: (ε, δ)-DP AS REGIONS
A mechanism M is $(\varepsilon, \delta)-$DP if and only if

$$\mathcal{R}(M) \subseteq \mathcal{R}(\varepsilon, \delta).$$

[5] discusses consequences of this view of differential privacy, notably the paper shows some easier proofs

of well known theorems regarding DP. One of them is the following.

PROPOSITION 9: DATA PROCESSING INEQUALITY FOR DP, AND CONVERSE
Let $M, M' : \mathcal{X}^n \to \mathcal{Y}$ be two mechanisms such that

$$X - M(X) - M'(X)$$

for all distributions of the dataset $X \in \mathcal{X}^n$ - we say that M **dominates** M'.
Then, for all $X_0, X_1 \in \mathcal{X}^n$ neighbouring,

$$\mathcal{R}\left(M', X_0, X_1\right) \subseteq \mathcal{R}\left(M, X_0, X_1\right).$$

Conversely, let $X_0, X_1 \in \mathcal{X}^n$ be neighbouring datasets. Let $Y \in \{M(X_0), M(X_1)\}$ and $Y' \in \{M'(X_0), M'(X_1)\}$ denote the random outputs of the mechanisms M and M'. If

$$\mathcal{R}\left(M', X_0, X_1\right) \subseteq \mathcal{R}\left(M, X_0, X_1\right).$$

then there exists a random mapping T such that

$$T(X) \sim Y.$$

2.3 WORST-CASE BOUNDS FOR COMPOSITION OF MECHANISMS

Section 2.1.4 discusses transforming the output of a mechanism M_1 through a function F and shows that the resulting mechanism $F \circ M_1$ is privacy-preserving. One may also wish to *compose* mechanisms in the following sense: given mechanisms $M_1, \ldots, M_L$ defined on a common set $\mathcal{X}^n$, one may want to output the ensemble $(M_1(X), \ldots, M_L(X))$ for the same input dataset $X \in \mathcal{X}^n$. This amounts to querying the dataset X multiple times, through multiple queries. The logic is the following: an adversary may be trying to uncover some information in X, and they may perform a first query $M_1(X)$, then select a second query $M_2(X; M_1(X))$ based on the result of the first query, then a third query $M_3(X; M_1(X), M_2(X, M_1(X)))$, etc. This is called **adaptive composition**.

Is this composition privacy-preserving, if the M_l's are privacy-preserving? How much information can the adversary gain as L increases, in the worst case?

2.3.1 BASIC COMPOSITION RESULT

The theorem below, taken from [5] or [3], gives the most direct answer to this question.

PROPOSITION 10: BASIC COMPOSITION RESULT
Let $M_1, \dots, M_L$ be mechanisms with independent internal randomness defined over the same set $\mathcal{X}^n$ with output in $\mathcal{Y}$ such that M_l is $(\varepsilon_l, \delta_l)-$DP. Define

$$\begin{aligned} M : \mathcal{X}^n &\to \mathcal{Y}^L \\ X &\mapsto M(X) = (M_1(X), \dots, M_L(X)). \end{aligned}$$

Then

$$M \text{ is } \left(\sum_{l=1}^{L} \varepsilon_l, \sum_{l=1}^{L} \delta_l \right) - \text{DP}.$$

Note that the assumption that the M_l's output in the same set $\mathcal{Y}$ is w.l.o.g as one can pick $\mathcal{Y}$ to be the union of the L output sets.
PROOF: It suffices to prove this proposition for $L = 2$. Let $X, X' \in \mathcal{X}^n$ be neighbouring datasets and let $T = T_1 \times T_2 \subseteq \mathcal{Y}^2$. Then

$$\begin{aligned} \mathbb{P}\left(M(X) \in T\right) &= \mathbb{P}\left(M_1(X) \in T_1, M_2(X) \in T_2\right) \\ &\overset{\text{indep.}}{=} \mathbb{P}\left(M_1(X) \in T_1\right) \mathbb{P}\left(M_2(X) \in T_2\right) \\ &\leq \left(e^{\varepsilon_1} \mathbb{P}\left(M_1(X') \in T_1\right) + \delta_1\right) \mathbb{P}\left(M_2(X) \in T_2\right) \\ &= e^{\varepsilon_1} \mathbb{P}\left(M_1(X') \in T_1\right) \mathbb{P}\left(M_2(X) \in T_2\right) + \delta_1 \underbrace{\mathbb{P}\left(M_2(X) \in T_2\right)}_{\leq 1} \\ &\leq e^{\varepsilon_1} \mathbb{P}\left(M_1(X') \in T_1\right) \mathbb{P}\left(M_2(X) \in T_2\right) + \delta_1 \\ &\leq e^{\varepsilon_1} \mathbb{P}\left(M_1(X') \in T_1\right) \left(e^{\varepsilon_2} \mathbb{P}\left(M_2(X') \in T_2\right) + \delta_2\right) + \delta_1 \\ &\leq e^{\varepsilon_1} \mathbb{P}\left(M_1(X') \in T_1\right) e^{\varepsilon_2} \mathbb{P}\left(M_2(X') \in T_2\right) + \delta_2 + \delta_1 \\ &= e^{\varepsilon_1 + \varepsilon_2} \mathbb{P}\left(M_1(X') \in T_1\right) \mathbb{P}\left(M_2(X') \in T_2\right) + \delta_1 + \delta_2 \\ &\overset{\text{indep.}}{=} e^{\varepsilon_1 + \varepsilon_2} \mathbb{P}\left(M_1(X') \in T_1, M_2(X') \in T_2\right) + \delta_1 + \delta_2 \\ &= e^{\varepsilon_1 + \varepsilon_2} \mathbb{P}\left(M(X') \in T\right) + \delta_1 + \delta_2. \end{aligned}$$

Then the proof for the general case follows by repeated application of the case for $L = 2$, since $M = (M_1, M_2, \dots, M_L)$ can be seen as $M = (M_1, \tilde{M}_2)$ with $\tilde{M}_2 = (M_2, \dots, M_L)$. □

If all ε_l and δ_l are equal, then M is $(k\varepsilon, k\delta)$ $-$DP, i.e the cost in ε and δ is **linear** in the number of mechanisms, in this basic setting. The proof is simple, but can we do better? The answer is *no*, if we do not give up anything. However, if we allow some slackness over the δ parameter...

2.3.2 OPTIMAL COMPOSITION RESULT

The theorem below is the main result of [5].

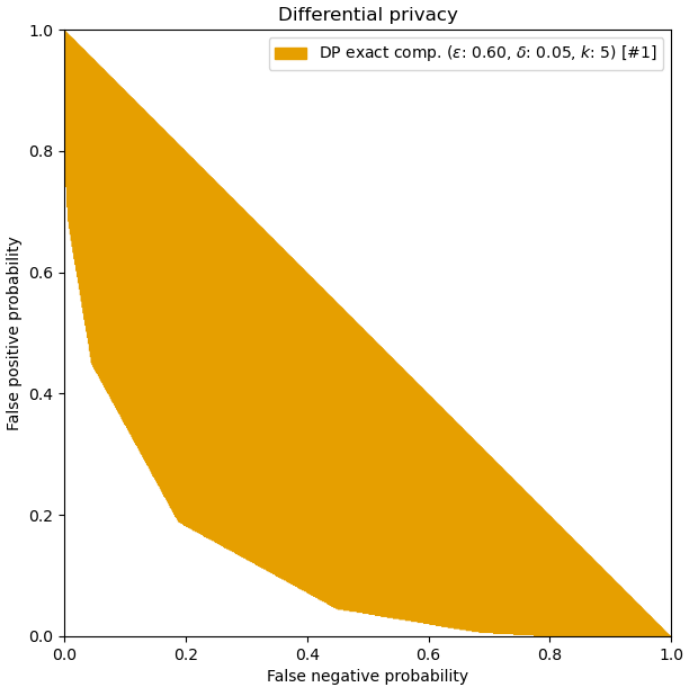

FIGURE 2.2
Region for 5-fold composition of (0.6, 0.05)-DP mechanisms

THEOREM 11: TIGHT WORST-CASE COMPOSITION RESULT
Let $M_1, \ldots, M_k$ be $(\varepsilon, \delta)-$DP mechanisms with independent internal randomness. Then their adaptive composition

$$M = (M_1, \ldots, M_k) \text{ is } \left((k-2i)\,\varepsilon, 1-(1-\delta)^k(1-\delta_i)\right) - \text{DP}$$

for $i \in [\![0, \lfloor \frac{k}{2} \rfloor]\!]$, where

$$\delta_i = \frac{\sum_{l=0}^{i-1} \binom{k}{l} \left(e^{(k-l)\varepsilon} - e^{(k-2i+l)\varepsilon}\right)}{(1+e^{\varepsilon})^k}.$$

This result tells us that the privacy region of a $k-$fold (ε, δ) composition is the **intersection** of

$$\left\{\mathcal{R}\left((k-2i)\varepsilon, 1-(1-\delta)^k(1-\delta_i)\right)\right\}_{i=1}^{\lfloor \frac{k}{2} \rfloor}.$$

[5] proves this theorem by constructing an explicit mechanism $M_{\varepsilon,\delta}$ then relying on the hypothesis testing interpretation of DP: The authors prove that $\mathcal{R}\left(M_{\varepsilon,\delta}\right) = \mathcal{R}\left(\varepsilon, \delta\right)$, thus showing that $M_{\varepsilon,\delta}$ dominates all $(\varepsilon, \delta)-$DP mechanisms. Lastly, they compute the privacy region of the k-fold composition of $M_{\varepsilon,\delta}$ with itself, yielding the result.

2.3.3 SIMPLIFIED COMPOSITION RESULT

In many applications, a closed form for the ε, δ parameters of the composition mechanism is desirable. The theorem below is also proven in [5], where some $\tilde{\delta}$ excess privacy budget is given.

THEOREM 12: SIMPLIFIED COMPOSITION RESULT
Let $M_1, ..., M_k$ be $(\varepsilon_l, \delta_l)-$DP with independent internal randomness where $l \in [\![1, k]\!]$. Also let $\tilde{\delta} > 0$. Then their adaptive composition

$$M = (M_1, \ldots, M_k) \text{ is } \left(\varepsilon_{\tilde{\delta}}, 1 - (1 - \tilde{\delta}) \prod_{l=1}^{k} (1 - \delta_l)\right) - \text{DP}$$

where

$$\varepsilon_{\tilde{\delta}} = \min \begin{cases} \sum_{l=1}^{k} \varepsilon_l, \\ \sum_{l=1}^{k} \frac{(e^{\varepsilon_l} - 1)\varepsilon_l}{e^{\varepsilon_l} + 1} + \sqrt{\sum_{l=1}^{k} 2\varepsilon_l^2 \log\left(\frac{1}{\tilde{\delta}}\right)}, \\ \sum_{l=1}^{k} \frac{(e^{\varepsilon_l} - 1)\varepsilon_l}{e^{\varepsilon_l} + 1} + \sqrt{\sum_{l=1}^{k} 2\varepsilon_l^2 \log\left(e + \frac{\sqrt{\sum_{l=1}^{k} \varepsilon_l^2}}{\tilde{\delta}}\right)}. \end{cases}$$

When all ε_l are equal to ε, these expressions simplify to

$$\varepsilon_{\tilde{\delta}} = \min \begin{cases} k\varepsilon, \\ \frac{(e^{\varepsilon} - 1)k\varepsilon}{e^{\varepsilon} + 1} + \varepsilon\sqrt{2k \log\left(\frac{1}{\tilde{\delta}}\right)}, \\ \frac{(e^{\varepsilon} - 1)k\varepsilon}{e^{\varepsilon} + 1} + \varepsilon\sqrt{2k \log\left(e + \frac{\sqrt{k\varepsilon^2}}{\tilde{\delta}}\right)}. \end{cases}$$

A comment is in order here. These bounds are worst case bounds on the differential privacy of composed mechanisms. Mechanisms that are commonly used, including those in 2.4, may achieve better DP rates under adaptive composition, especially when considering sensible generalisations of DP, discussed in 2.5.

2.4 COMMON MECHANISMS FOR DIFFERENTIAL PRIVACY

This section examines commonly used mechanisms and their DP guarantees.

2.4.1 SENSITIVITY

We start by defining the **sensitivity** of a function $f : \mathcal{X}^n \to \mathcal{Y}$ as done in section 3.3 of [3].

DEFINITION 13: METRIC
Let S be a set, and $d : S \times S \to \mathbb{R}$. d is a **metric** or **distance function** if and only if, for all $a, b, c \in S$, all three of the following properties hold:

- Positivity: $d(a) \geq 0$,
- Symmetry: $d(a, b) = d(b, a)$,
- Triangle inequality: $d(a, b) \leq d(a, c) + d(c, b)$.

DEFINITION 14: d-SENSITIVITY
Let d be a metric on $\mathcal{Y}$. Let $f : \mathcal{X}^n \to \mathcal{Y}$. The $d-$**sensitivity of** f is defined by

$$\Delta_{f,d} = \sup_{\substack{X,X' \in \mathcal{X}^n \\ X,X' \text{ neighbouring}}} d\left(f(X), f(X')\right).$$

We will drop the subscripts f, d when the function or the metric are clear from context. In the following, the ℓ_1 and ℓ_2 sensitivities will be discussed.

2.4.2 LAPLACE MECHANISM

The Laplace mechanism can achieve $(\varepsilon, 0)$-DP for any numerical query. Firstly, we recall the definition of the Laplace distribution.

DEFINITION 15: LAPLACE DISTRIBUTION
Let $b > 0$. A random variable X follows the Laplace (μ, b) distribution when X has the density

$$P_X(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right).$$

Note that $\mathbb{E}(X) = \mu$ and $\text{Var}(X) = 2b^2$.

We can now construct the Laplace mechanism then prove its privacy-preserving property.

DEFINITION 16: LAPLACE MECHANISM
Let $\varepsilon > 0$ and $f : \mathcal{X}^n \to \mathbb{R}^k$ with ℓ_1-sensitivity Δ. The **Laplace mechanism** is defined as

$$\begin{aligned} M : \mathcal{X}^n &\to \mathbb{R}^k \\ X &\mapsto M(X) = f(X) + (Y_1, \ldots, Y_k) \end{aligned}$$

where $\{Y_i\}_{i=1}^k \overset{\text{i.i.d}}{\sim} \text{Laplace}\left(0, \frac{\Delta}{\varepsilon}\right)$.

PROPOSITION 17: PRIVACY OF THE LAPLACE MECHANISM
The Laplace mechanism is $(\varepsilon, 0)-$DP.

PROOF: Let X_1, X_2 be neighbouring datasets in $\mathcal{X}^n$. Observe that $M(X_i) \sim \text{Laplace}\left(f(X_i), \frac{\Delta}{\varepsilon}\right)$. Let $y \in \mathbb{R}^k$.

$$\frac{P_{M(X_1)}(y)}{P_{M(X_2)}(y)} = \prod_{i=1}^{k} \frac{\exp\left(-\frac{\varepsilon}{\Delta}|f(X_1)_i - y_i|\right)}{\exp\left(-\frac{\varepsilon}{\Delta}|f(X_2)_i - y_i|\right)}$$

$$
\begin{aligned}
&= \prod_{i=1}^{k} \exp\left[-\frac{\varepsilon}{\Delta}\left(|f(X_1)_i - y_i| - |f(X_2)_i - y_i|\right)\right] \\
&\leq \prod_{i=1}^{k} \exp\left[\frac{\varepsilon}{\Delta}|f(X_1)_i - f(X_2)_i|\right] \text{ (Triangle inequality)} \\
&= \exp\left[\varepsilon \frac{\sum_{i=1}^{k}|f(X_1)_i - f(X_2)_i|}{\Delta}\right] \\
&= \exp\left[\frac{\varepsilon}{\Delta}\ell_1\left(f(X_1), f(X_2)\right)\right] \\
&\leq \exp(\varepsilon)\ (\ell_1 - \text{sensitivity of } f).
\end{aligned}
$$

Thus $\mathbb{P}\left(M(X_1) \in T\right) \leq e^{\varepsilon}\mathbb{P}\left(M(X_2) \in T\right)$ for any measurable set $T \subseteq \mathbb{R}^k$ follows easily. □

2.4.3 GAUSSIAN MECHANISM

The Laplace mechanism is an $(\varepsilon, 0)$ −DP mechanism, here we introduce an (ε, δ) −DP mechanism for numerical queries, for $\delta > 0$.

DEFINITION 18: GAUSSIAN MECHANISM
Let $\varepsilon > 0$, $\delta\ ]0, 1]$, and $f : \mathcal{X}^n \to \mathbb{R}^k$ with ℓ_2-sensitivity Δ. The **Gaussian mechanism** is defined as

$$
\begin{aligned}
M : \mathcal{X}^n &\to \mathbb{R}^k \\
X &\mapsto M(X) = f(X) + (Y_1, \ldots, Y_k)
\end{aligned}
$$

where $\{Y_i\}_{i=1}^{k} \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(0, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right)$.

PROPOSITION 19: PRIVACY OF THE GAUSSIAN MECHANISM
The Gaussian mechanism is $(\varepsilon, \delta)-$DP.

A proof of this proposition can be found in the appendix of [6], a sketch of it in lecture 5 of [2].

2.4.4 GENERALIZED RANDOMIZED RESPONSE

As a follow-up to the binary randomized response mechanism presented in 2.1.2, we introduce a generalized notion of **randomized response**.

DEFINITION 20: GENERALIZED RANDOMIZED RESPONSE
Let $\varepsilon \geq 0$ and let $\mathcal{X}$ be a finite set. The **randomized response mechanism** is defined as

$$\begin{aligned} M : \mathcal{X}^n &\to \mathcal{X}^n \\ X &\mapsto M(X) = (M(X)_i)_{i=1}^n \text{ where} \end{aligned}$$

$$M(X)_i = \begin{cases} X_i & \text{w.p } p_\varepsilon, \\ \text{Unif}(\mathcal{X}) & \text{w.p } 1 - p_\varepsilon \end{cases} \text{ with } p_\varepsilon = \frac{e^\varepsilon - 1}{e^\varepsilon + |\mathcal{X}| - 1}.$$

We then prove that it is privacy-preserving.

PROPOSITION 21: PRIVACY OF GENERALIZED RANDOMIZED RESPONSE
The randomized response mechanism is $(\varepsilon, 0)-$DP.

PROOF: First, observe the following. By the law of total probability, we find

$$\mathbb{P}\left(M(X)_i = x\right) = \begin{cases} p_\varepsilon + \frac{1-p_\varepsilon}{|\mathcal{X}|} & \text{if } X_i = x, \\ \frac{1-p_\varepsilon}{|\mathcal{X}|} & \text{if } X_i \neq x. \end{cases}$$

Let $X_0, X_1 \in \mathcal{X}^n$ be two neighbouring datasets. W.l.o.g we assume they differ in their first position. Thus, for any $x^n \in \mathcal{X}^n$,

$$\begin{aligned} \frac{\mathbb{P}\left(M(X_0) = x^n\right)}{\mathbb{P}\left(M(X_1) = x^n\right)} &= \frac{\mathbb{P}\left(M(X_0)_1 = x_1\right)}{\mathbb{P}\left(M(X_1)_1 = x_1\right)} \\ &= \begin{cases} \frac{p_\varepsilon + \frac{1-p_\varepsilon}{|\mathcal{X}|}}{\frac{1-p_\varepsilon}{|\mathcal{X}|}} & \text{if } x_1 = X_{0,1} \neq X_{1,1}, \\ \frac{\frac{1-p_\varepsilon}{|\mathcal{X}|}}{p_\varepsilon + \frac{1-p_\varepsilon}{|\mathcal{X}|}} & \text{if } x_1 = X_{1,1} \neq X_{0,1}, \\ 1 & \text{if } x_1 \neq X_{0,1} \text{ and } x_1 \neq X_{1,1}. \end{cases} \end{aligned}$$

Thus it tightly holds that

$$\frac{\mathbb{P}\left(M(X_0) = x^n\right)}{\mathbb{P}\left(M(X_1) = x^n\right)} \leq \frac{p_\varepsilon + \frac{1-p_\varepsilon}{|\mathcal{X}|}}{\frac{1-p_\varepsilon}{|\mathcal{X}|}}.$$

Consequently, for the mechanism to be $(\varepsilon, 0)$-DP, we select p_ε so that

$$\frac{p_\varepsilon + \frac{1-p_\varepsilon}{|\mathcal{X}|}}{\frac{1-p_\varepsilon}{|\mathcal{X}|}} = e^\varepsilon.$$

Solving for p_ε:

$$\begin{aligned} \frac{p_\varepsilon + \frac{1-p_\varepsilon}{|\mathcal{X}|}}{\frac{1-p_\varepsilon}{|\mathcal{X}|}} = e^\varepsilon &\iff p_\varepsilon + \frac{1}{|\mathcal{X}|} - \frac{p_\varepsilon}{|\mathcal{X}|} = \frac{e^\varepsilon}{|\mathcal{X}|} - \frac{e^\varepsilon}{|\mathcal{X}|} p_\varepsilon \\ &\iff p_\varepsilon \left(1 - \frac{1}{|\mathcal{X}|}\right) + \frac{e^\varepsilon}{|\mathcal{X}|} p_\varepsilon = \frac{e^\varepsilon - 1}{|\mathcal{X}|} \end{aligned}$$

$$\iff p_\varepsilon(|\mathcal{X}| - 1 + e^\varepsilon) = e^\varepsilon - 1$$
$$\iff p_\varepsilon = \frac{e^\varepsilon - 1}{e^\varepsilon + |\mathcal{X}| - 1}.$$

□

2.4.5 EXPONENTIAL MECHANISM

A popular and useful mechanism for optimisation of a score function under privacy constraints is the **exponential mechanism**, as it features an embedded score function in its definition.

DEFINITION 22: EXPONENTIAL MECHANISM
Let $\varepsilon \geq 0$ and let $\mathcal{X}$ be a set. Denote by $\mathcal{C}$ a finite set, and let $s : \mathcal{X}^n \times \mathcal{C} \to \mathbb{R}$ be a **score function** such that $s(\cdot, c)$ has ℓ_1-sensitivity upper bounded by Δ for all $c \in \mathcal{C}$. The **exponential mechanism** is defined as

$$M : \mathcal{X}^n \to \mathcal{C}$$
$$X \mapsto M(X) = c \in \mathcal{C}$$

such that

$$\mathbb{P}\left(M(X) = c\right) = \frac{\exp\left(\frac{\varepsilon s(X,c)}{2\Delta}\right)}{\sum_{c' \in \mathcal{C}} \exp\left(\frac{\varepsilon s(X,c')}{2\Delta}\right)}.$$

Note that the definition naturally extends to the general case where $\mathcal{C}$ is not finite.

It is relatively straightforward to check its privacy-preserving property. The proof is similar to the Laplace mechanism's and can be found in B.1.2.

PROPOSITION 23: PRIVACY OF EXPONENTIAL MECHANISM
The exponential mechanism is $(\varepsilon, 0)-$DP.

The exponential mechanism is interesting as it has an embedded utility. We can say more about it.

PROPOSITION 24: UTILITY GUARANTEE FOR THE EXPONENTIAL MECHANISM
Assume $\mathcal{C}$ is finite. Let $X \in \mathcal{X}^n$ and define $\text{OPT}(X) = \max_{c \in \mathcal{C}} s(X, c)$. Let $\mathcal{C}^* = \{c \in \mathcal{C} \mid s(X, c) = \text{OPT}(X)\}$. Then

$$\mathbb{P}\left(s(X, M(X)) \leq \text{OPT}(X) - \frac{2\Delta}{\varepsilon}\left(\log\left(\frac{|\mathcal{C}|}{|\mathcal{C}^*|}\right) + t\right)\right) \leq \exp(-t).$$

Consequently,

$$\mathbb{P}\left(s(X, M(X)) \leq \text{OPT}(X) - \frac{2\Delta}{\varepsilon}\left(\log\left(|\mathcal{C}|\right) + t\right)\right) \leq \exp(-t).$$

PROOF: Let $c \in \mathcal{C}$. Then

$$\begin{aligned}
\mathbb{P}\left(s(X, M(X)) \leq t\right) &\leq \frac{\sum_{\{c \in \mathcal{C} \mid s(X,c) \leq t\}} \exp\left(\frac{\varepsilon s(X,c)}{2\Delta}\right)}{\sum_{c' \in \mathcal{C}} \exp\left(\frac{\varepsilon s(X,c')}{2\Delta}\right)} \\
&\leq \frac{|\mathcal{C}| \exp\left(\frac{\varepsilon t}{2\Delta}\right)}{|\mathcal{C}^*| \exp\left(\frac{\varepsilon \mathrm{OPT}(X)}{2\Delta}\right)} \text{ (larger numerator, smaller denominator)} \\
&= \frac{|\mathcal{C}|}{|\mathcal{C}^*|} \exp\left(\frac{\varepsilon}{2\Delta}(t - \mathrm{OPT}(X))\right).
\end{aligned}$$

□

Observe that the upper bound on the second line is quite loose. Yet, we will see later that this bound still proves meaningful in some cases.

2.5 VARIATIONS OF DIFFERENTIAL PRIVACY

As we have seen in section 2.1.4, the worst case bounds for the composition of differentially private mechanisms are pessimistic; in the sense that, despite being tight as there exists setups that reach these bounds, they are worse than the real privacy guarantees one can obtain for commonly used mechanisms. Consequentially, many authors have proposed variations and generalisations of differential privacy, to get more pragmatically interesting bounds, as well as relax or modify the underlying threat model implied by differential privacy. Some examples of these variations and generalisations include, but are by no means limited to:

- Rényi differential privacy [7]: $(\varepsilon, 0)-$DP is equivalent to the following condition on the Rényi divergence of order infinity of the outputs' distributions, $\mathrm{D}_\infty\left(P_{M(X_0)} || P_{M(X_1)}\right) \leq \varepsilon$. From this perspective, a natural extension is to substitute D_∞ for D_α with $\alpha > 1$.
- local differential privacy [8]: in the LDP setup, we consider that all datasets of $\mathcal{X}^n$ are neighbouring.
- DP with total variation [9].
- $f-$DP and Gaussian DP [10].

In this section, we discuss the latter two in some detail.

2.5.1 TOTAL VARIATION WITH DIFFERENTIAL PRIVACY

The concept of $(\varepsilon, \delta)-$DP with $\eta-$TV was introduced in [9]. Its privacy regions tend to be much closer to the exact privacy regions of some popular mechanisms, e.g Laplace, than the worst case bounds defined above. This is achieved by considering a natural notion of the total variation of a mechanism, i.e a measure of how much the distribution of its output changes when altering one element of the dataset.

We start by defining the **trade-off function** or **ROC** (receiver operating characteristic).

DEFINITION 25: TRADE-OFF FUNCTION, ROC
Let H_0 and H_1 be the two hypotheses of a binary hypothesis test with data $S \in \mathcal{S}$ such that $S \sim P_i$ under hypothesis H_i, $i \in \{0, 1\}$. Let ϕ be a possibly randomized decision function $\phi : \mathcal{S} \to \{0, 1\}$. Then the **trade-off function** or **ROC curve** (receiver operating characteristic) is defined by

$$T(P_0, P_1)(\alpha) = \inf_{\phi:\, \mathcal{S}\to\{0,1\}} \{\mathbb{P}(\phi(S) = 0 \mid S \sim P_1) \mid \mathbb{P}(\phi(S) = 1 \mid S \sim P_0) \leq \alpha\}.$$

Graphically, the ROC curve is the bottom boundary of the (convex) privacy region.

We also need the definition of **total variation**.

DEFINITION 26: TOTAL VARIATION
Let P and Q be two distributions over the alphabet $\mathcal{S}$. Then their **total variation** is defined by

$$d_{TV}(P, Q) = \sup_{A\subseteq\mathcal{S}} \{P(A) - Q(A)\}.$$

For an observed $Y \in \{M(X_0), M(X_1)\}$ for $X_0, X_1 \in \mathcal{X}^n$, we denote by P_i the distribution of $M(X_i)$. Thus we can define the **total variation of a mechanism**.

DEFINITION 27: TOTAL VARIATION OF A MECHANISM
Given $\eta \in [0, 1]$, a mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ has **total variation less than** η, i.e

$$d_{TV}(M) \leq \eta, \text{ or } M \text{ is } \eta - TV,$$

if for all neighbouring $X_0, X_1 \in \mathcal{X}^n$,

$$d_{TV}(P_0, P_1) \leq \eta.$$

We can characterize this new property as such.

COROLLARY 28: CHARACTERIZATION OF η-TV
Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a mechanism.

$$M \text{ is } \eta - TV$$
$$\iff$$
$$M \text{ is } (0, \eta) - \text{DP}.$$

This implies that M is (ε, δ) $-$DP and η-TV if and only if for all neighbouring datasets $X_0, X_1 \in \mathcal{X}^n$

$$\begin{aligned} \alpha + e^{\varepsilon} T(P_0, P_1)(\alpha) &\geq 1 - \delta, \\ e^{\varepsilon}\alpha + T(P_0, P_1)(\alpha) &\geq 1 - \delta \text{ and} \\ \alpha + T(P_0, P_1)(\alpha) &\geq 1 - \eta. \end{aligned}$$

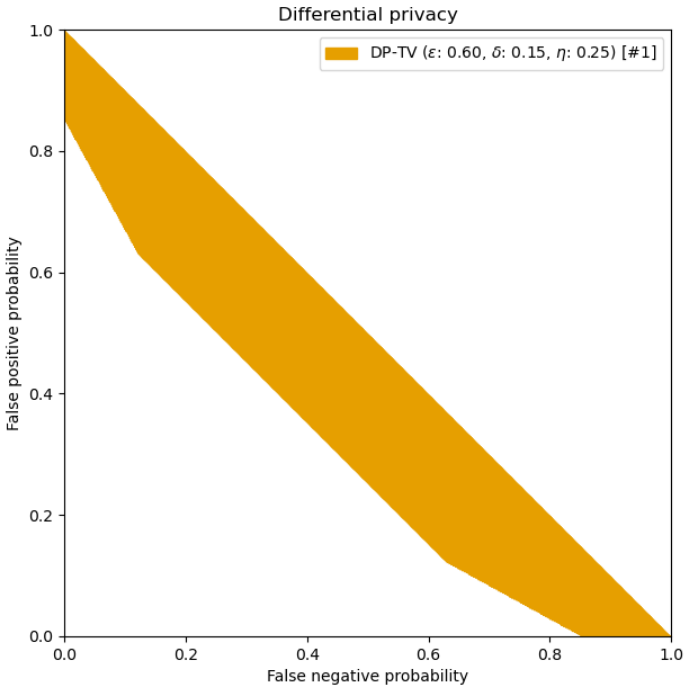

FIGURE 2.3
(0.6, 0.15)-DP and 0.25-TV region

It is clear that all $(\varepsilon, \delta)-$DP mechanisms are $1-$TV. Hence this new notion of privacy is indeed a generalization of DP. We will often refer to it as DP-TV.

Keeping track of the total variation gives the following theorem, that yields privacy guarantees which are much closer to reality for common mechanisms.

THEOREM 29: COMPOSITION RESULT, WITH TV
Let $M_1, \dots, M_k$ be $(\varepsilon, \delta)-$DP and $\eta-$TV mechanisms with independent internal randomness where $\varepsilon \geq 0$, $\delta \in [0, 1]$ and $\eta \in \left[\delta, \delta + (1-\delta)\frac{e^\varepsilon - 1}{e^\varepsilon + 1}\right]$. Then their adaptive composition

$$M = (M_1, \dots, M_k) \text{ is } \left(j\varepsilon, 1 - (1-\delta)^k(1-\delta_j)\right) - \text{DP and } \eta' - \text{TV}$$

for all $j \in [\![0, k]\!]$, where

$$\delta_j = \sum_{i=0}^{k-j-1} \binom{k}{i} \sum_{l=0}^{\lceil \frac{k-j-i}{2} \rceil - 1} \binom{k-i}{l} \left(\frac{\mu - 1}{e^\varepsilon + 1}\right)^{k-i} \mu^i \left(e^{(k-l-i)\varepsilon} - e^{(l+j)\varepsilon}\right),$$

$$\mu = 1 - \frac{\delta - \eta}{\delta - 1} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}, \text{ and}$$

$$\eta' = 1 - (1-\delta)^k(1-\delta_0).$$

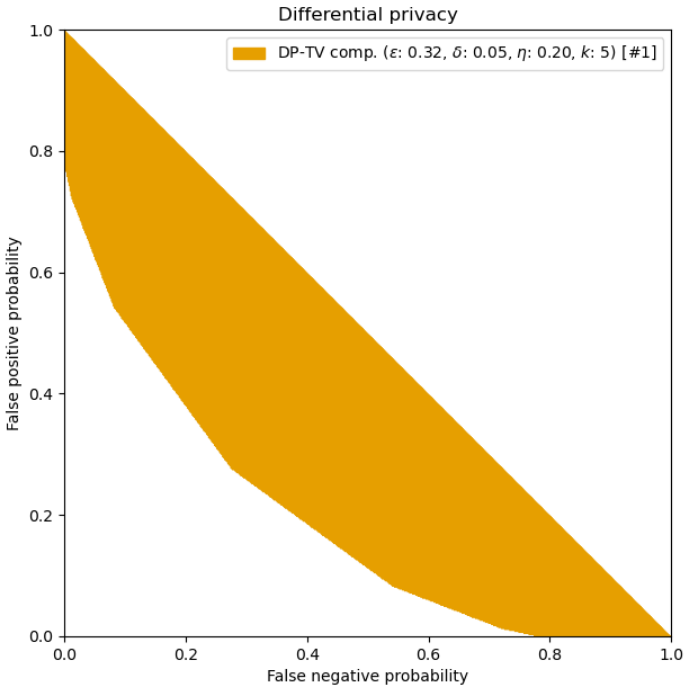

FIGURE 2.4
Region for 5-fold composition of (0.6,0.05)-DP and 0.2-TV mechanisms

The proof is given in [9], following the same machinery as the one developed in [5].

2.5.2 $f-$DP AND GAUSSIAN DIFFERENTIAL PRIVACY

$f-$DP is a relative notion of privacy, where the privacy of a mechanism is compared to a trade-off function f, i.e $f = T(P, Q)$ for some choice of distributions P and Q. [10] shows that it is in many common cases an interesting notion of privacy, including under composition, due to its algebraic nature. The following proposition from [10] characterizes such fs.

PROPOSITION 30: TRADE-OFF FUNCTIONS CHARACTERIZATION
A function $f : [0, 1] \to [0, 1]$ is a trade-off function $f = T(P, Q)$ for some choice of distributions P, Q if and only if

- f is convex,
- f is continuous,
- f is non-increasing, and
- $f(x) \leq 1 - x$ for $x \in [0, 1]$.

We will say f is **symmetric** when $f = f^{-1}$, i.e the curve of f is axially symmetric with respect to the line $y = x$.

Thus we can define $f-$DP.

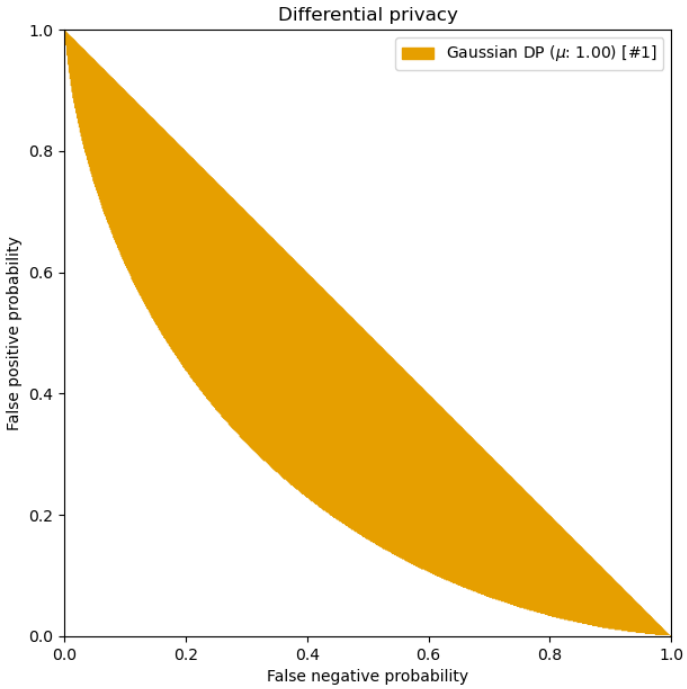

FIGURE 2.5
1-GDP region

DEFINITION 31: $f-$DP AND GAUSSIAN DP
Let f be a trade-off function $f = T(P, Q)$ for some distributions P and Q. A mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ is $f-$**DP** if, for all neighbouring $X_0, X_1 \in \mathcal{X}^n$

$$T(P_0, P_1) \geq f.$$

In particular, if $f = G_\mu = T(\mathcal{N}(0,1), \mathcal{N}(\mu,1))$, then we say that M is $\mu-$**Gaussian differentially private**, or M is $\mu-$**GDP**.

Note that $G_\mu(\alpha) = \Phi(\Phi^{-1}(1-\alpha) - \mu)$ where Φ is the CDF of $Z \sim \mathcal{N}(0,1)$. Intuitively, M being $\mu-$GDP implies that guessing $i \in \{0,1\}$ from M's output $Y \in \{M(X_0), M(X_1)\}$ is as hard as distinguishing $\mathcal{N}(0,1)$ from $\mathcal{N}(\mu,1)$. The smaller μ, the better the privacy guarantee.

Also, this relative definition of privacy is indeed a generalisation of $(\varepsilon, \delta)-$DP as the following corollary from [10] observes.

COROLLARY 32: (ε, δ) $-$DP IS $f-$DP FOR SOME f
Let $f_{\varepsilon,\delta}(\alpha) = \max\{0, 1-\delta-e^\varepsilon\alpha, e^{-\varepsilon}(1-\delta-\alpha)\}$. Then M is (ε, δ) $-$DP if and only if M is $f_{\varepsilon,\delta}-$DP.

Adaptive composition of $f-$DP mechanisms follows a simple **tensor product** algebraic rule.

DEFINITION 33: TENSOR PRODUCT OF TRADE-OFF FUNCTIONS
Let P, P', Q, Q' be probability distributions. Let $f = T(P, Q)$ and $g = T(P', Q')$. Then the **tensor product** of f and g is defined as

$$f \otimes g = T(P \times P', Q \times Q').$$

We will write the n fold tensor product of f with itself as $f^{\otimes n}$.

The proof that this operation is well defined can be found in [10]. Then we can show the following result on the composition of $f-$DP mechanisms.

THEOREM 34: COMPOSITION RESULT, FOR $f-$DP
Let $M_1, \ldots, M_k$ be f_i-DP mechanisms with independent internal randomness. Then their adaptive composition

$$M = (M_1, \ldots, M_k) \text{ is } \bigotimes_{i=1}^{k} f_i - \text{DP}.$$

This privacy guarantee is tight.

The proof can be found in [10]. In particular, the composition of μ_i-GDP is particularly simple.

COROLLARY 35: COMPOSITION RESULT, FOR $\mu-$GDP
Let $M_1, \ldots, M_k$ be μ_i-GDP mechanisms with independent internal randomness. Then the adaptive composition

$$M = (M_1, \ldots, M_k) \text{ is } \sqrt{\sum_{i=1}^{k} u_i^2} - \text{ GDP}.$$

This bound is tight.

PROOF: It suffices to prove that, for $k = 2$,

$$T\left(\mathcal{N}(0, I_2), \mathcal{N}\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, I_2\right)\right) = T\left(\mathcal{N}(0,1), \mathcal{N}\left(\sqrt{u_1^2 + u_2^2}, 1\right)\right).$$

Then we apply the previous identity repeatedly for a general $k \geq 1$. This identity can be seen as follows:

$$\begin{aligned}
&T\left(\mathcal{N}(0, I_2), \mathcal{N}\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, I_2\right)\right) \\
&= T\left(\mathcal{N}(0,1) \times \mathcal{N}(0,1), \mathcal{N}\left(\sqrt{\mu_1^2 + \mu_2^2}, 1\right) \times \mathcal{N}(0,1)\right) \quad \text{by rotational invariance of } \mathcal{N} \\
&= T\left(\mathcal{N}(0,1), \mathcal{N}\left(\sqrt{\mu_1^2 + \mu_2^2}, 1\right)\right) \otimes T\left(\mathcal{N}(0,1), \mathcal{N}(0,1)\right)
\end{aligned}$$

$$= G_{\sqrt{\mu_1^2+\mu_2^2}} \otimes \mathrm{id}$$
$$= G_{\sqrt{\mu_1^2+\mu_2^2}}.$$

□

What's more, the authors of [10] prove a very general central-limit like theorem on the composition of mechanisms. Some notation first, for a trade-off function $f = T(P, Q)$, denote the KL divergence of P and Q by

$$\mathrm{kl}(f) = \mathrm{D}(P||Q) = -\int_0^1 \log|f'(x)|\mathrm{d}x.$$

Furthermore, define

$$\kappa_2(f) = \int_0^1 \log^2\left|f'(x)\right|\mathrm{d}x,\ \kappa_3(f) = \int_0^1 \left|\log\left|f'(x)\right|\right|^3 \mathrm{d}x \text{ and } \bar{\kappa}_3(f) = \int_0^1 \left|\log\left|f'(x)\right| + \mathrm{kl}(f)\right|^3 \mathrm{d}x.$$

THEOREM 36: CENTRAL-LIMIT LIKE THEOREM FOR $f-$DP COMPOSITION

Let $f_1, \ldots, f_k$ be symmetric trade-off functions such that $\kappa_3(f_i) < +\infty$ for all $i \in [\![1, n]\!]$. Denote $\mathrm{kl} = (\mathrm{kl}(f_i))_{i=1}^n$, $\kappa_2 = (\kappa_2(f_i))_{i=1}^n$, $\kappa_3 = (\kappa_3(f_i))_{i=1}^n$, $\bar{\kappa}_3 = (\bar{\kappa}_3(f_i))_{i=1}^n$,

$$\mu = \frac{2\,||\mathrm{kl}||_1}{\sqrt{||\kappa_2||_1 - ||\mathrm{kl}||_2^2}} \text{ and } \gamma = \frac{0.56\,||\bar{\kappa}_3||_1}{\left(||\kappa_2||_1 - ||\mathrm{kl}||_2^2\right)^{\frac{3}{2}}}.$$

If $\gamma < \frac{1}{2}$. Then, for all $\alpha \in [\gamma, 1-\gamma]$,

$$G_\mu(\alpha+\gamma) - \gamma \leq \left(\bigotimes_{i=1}^n f_i\right)(\alpha) \leq G_\mu(\alpha-\gamma) + \gamma.$$

This theorem can be rewritten in an asymptotic form.

THEOREM 37: ASYMPTOTICS OF $f-$DP COMPOSITION

Let $\{\{f_{ni}\} \mid i \in [\![1, n]\!]\}_{n=1}^\infty$ be a triangular array of symmetric trade-off functions. Let $K \geq 0$, $s > 0$. Assume:

- $\sum_{i=1}^\infty \mathrm{kl}(f_{ni}) = K$,
- $\lim\limits_{n\to\infty} \max_{i\in[\![1,n]\!]} \mathrm{kl}(f_{ni}) = 0$,
- $\sum_{i=1}^\infty \kappa_2(f_{ni}) = s^2$, and
- $\sum_{i=1}^\infty \kappa_3(f_{ni}) = 0$.

Then

$$\lim_{n\to\infty} \bigotimes_{i=1}^n f_{ni} = G_{\frac{2K}{s}}$$

uniformly over $[0, 1]$.

In short, the adaptive composition of n privacy-preserving mechanisms is asymptotically $\frac{2K}{s}-$GDP.

CHAPTER 3

VISUALISING PRIVACY

In this chapter, we detail the features of the visualisation tool along with some implementation details, both mathematical and software oriented. The relevant code can be found at `https://github.com/Dicedead/privacyVis`.

3.1 OVERVIEW OF THE VISUALISATION TOOL

When starting the software, a main menu prompts a choice between the two following types of windows:

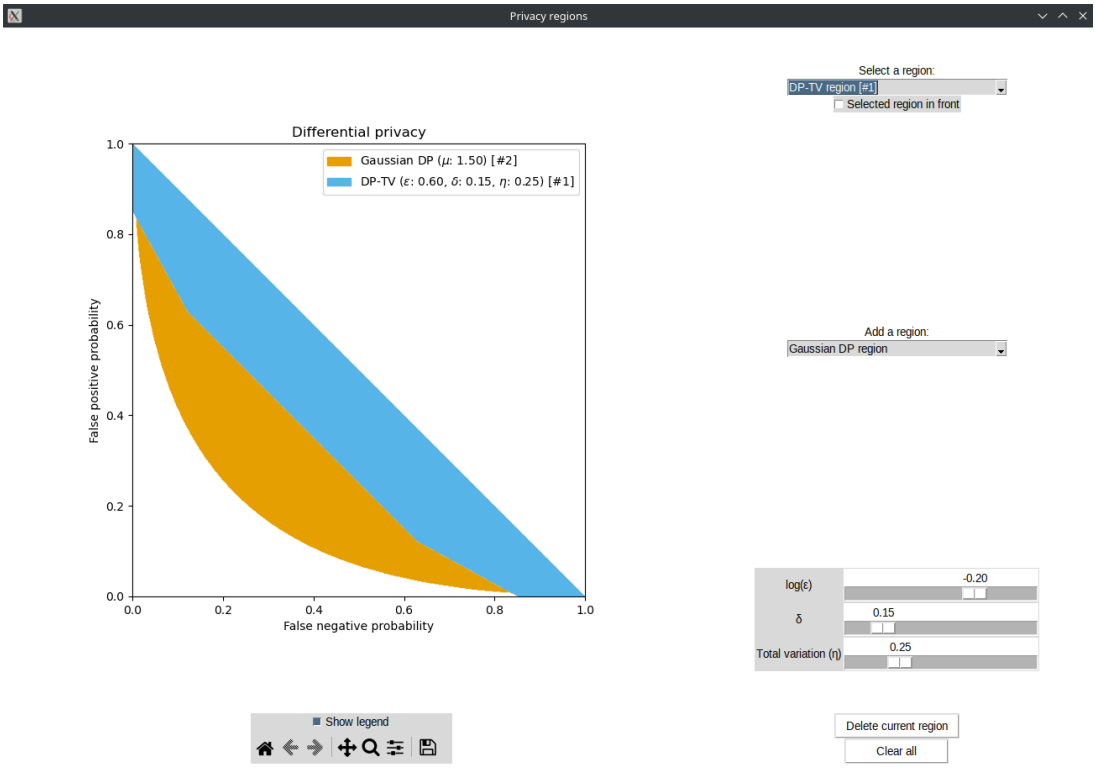

FIGURE 3.1
Privacy regions window

- A privacy regions window as depicted in figure 3.1. It showcases regions of previously listed

mechanisms as well as regions corresponding to DP, its variations and some composition results.

- A utility-privacy trade-off window as depicted in figure 3.2 for a specific combination of query, utility and mechanism. It shows how the selected utility evolves as a function of the parameters of the privacy-preserving mechanism, along with the corresponding privacy region.

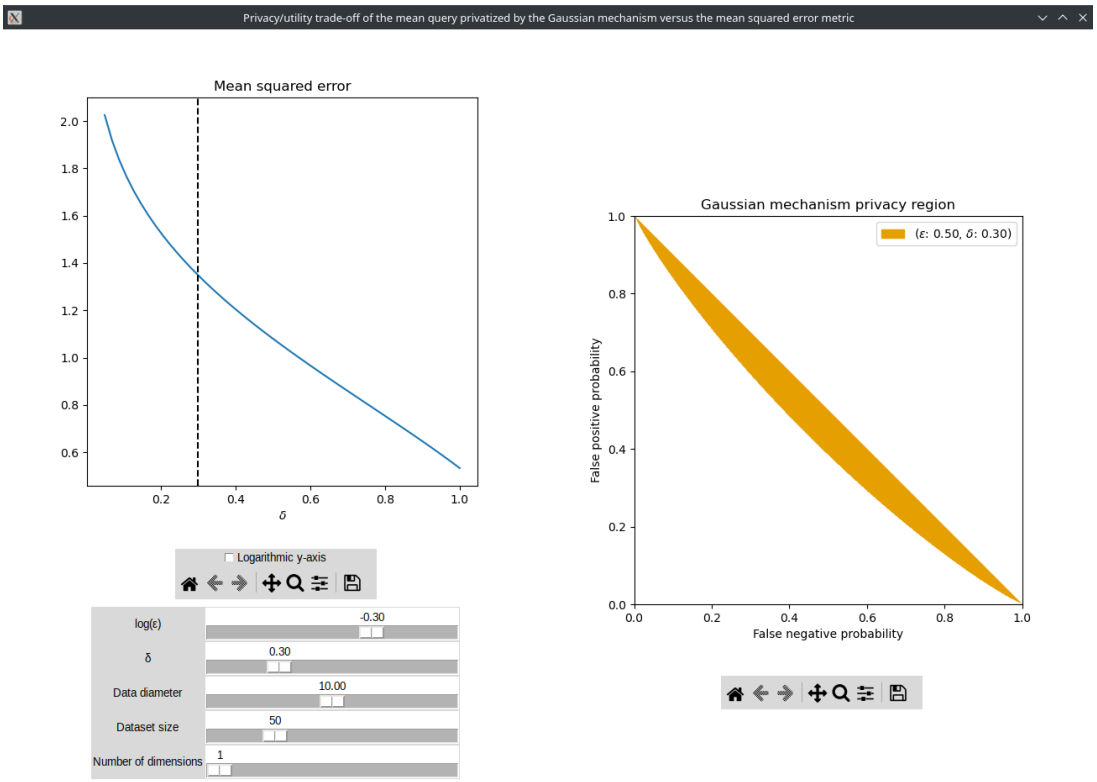

FIGURE 3.2
Mean privacy-utility window

3.2 SOFTWARE ARCHITECTURE

We explain some of the fundamental software concepts that went into the making of this tool.

3.2.1 REPRESENTING PRIVACY REGIONS

A privacy region $\mathcal{R}$ is a subset of $[0,1]^2$. As a reminder, the privacy region for (vanilla) $(\varepsilon, \delta)-$DP is

$$\mathcal{R}(\varepsilon,\delta) = \left\{ (\mathrm{FN},\mathrm{FP}) \in [0,1]^2 \mid \begin{cases} \mathrm{FP} + e^{\varepsilon}\mathrm{FN} \geq 1-\delta, \\ e^{\varepsilon}\mathrm{FP} + \mathrm{FN} \geq 1-\delta. \end{cases} \right\}.$$

We needed to find a way to represent regions with low memory and time complexities, and independently from their graphical representation for increased modularity. An intuitive option is to represent a region $\mathcal{R}$ by a *two-dimensional boolean array* R, representing a mesh on $[0,1]^2$. If $R[x,y]$ is set to 1 then the pixel at coordinates (x,y) is inside the region $\mathcal{R}$, else it is not. While this approach is easy and allows for relatively simple post-processing, e.g intersection of regions, its memory burden is too high.

Hence, we have opted for a *functional* approach, just one layer of abstraction higher: a `Region` is described by a sequence of constraints, of type `Sequence[Constraint]` where each `Constraint`

is a predicate on $[0,1]^2$:

$$p \in \texttt{Constraint} \implies p : [0,1]^2 \to \{0,1\}\,.$$

For instance, $\mathcal{R}(\varepsilon, \delta)$ is represented by a list of two constraints $[p_1, p_2]$:

- $p_1(\text{FP}, \text{FN}) = \mathbb{1}\{\text{FP} + e^{\varepsilon}\text{FN} \geq 1 - \delta\}$,
- $p_2(\text{FP}, \text{FN}) = \mathbb{1}\{e^{\varepsilon}\text{FP} + \text{FN} \geq 1 - \delta\}$.

In practice, we also add a constraint $p_3(\text{FP}, \text{FN}) = \mathbb{1}\{\text{FP} + \text{FN} \leq 1\}$. $\text{FP} + \text{FN} = 1$ is the diagonal line connecting $(1, 0)$ to $(0, 1)$ corresponding to strategies picking one of the two hypotheses at random. Hence why we want to plot only the portion of the privacy region that stays under this line.

Since functions are cheap to store, this symbolic approach is very economical in memory and is as malleable as the naive solution explained above. Indeed, *intersecting* regions $\mathcal{R}_1, \ldots, \mathcal{R}_m$ is simply done by concatenating their m sequences of constraints into a single list, defining $\cap_{i=1}^{m} \mathcal{R}_i$.

Most regions of interest are defined in `model/diff_privacy/regions.py`.

Note that regions could also have been represented by their bottom boundary, the ROC curve or `TradeOffFunction` as found in the code $f(\text{FP})$. The equivalent constraint map is

$$p_f(\text{FP}, \text{FN}) = \mathbb{1}\{\text{FN} \geq f(\text{FP})\}\,.$$

We chose not to represent regions by their sole ROC bottom boundary, even though it is in practice the only characteristic that we actually care about, to adhere to the existing literature's region representation (see figures in [5], [9]).

3.2.2 DRAWING PRIVACY REGIONS

Region drawing is handled in `gui/region_figures.py`, in the class `MultiRegionFigure`. The mesh is initialised in the class constructor, then the user may add regions through the `add_region` method. Then, `draw_figure` applies the constraints of each region, reorders the computed regions - in the way that will be shortly explained - and lastly displays all regions on the same plot.

About reordering. The rationale is to push larger regions towards the back, so smaller regions at the front are more visible. Thus, given two computed regions $\{R_1[x,y]\}_{x \in [\![1,N]\!], y \in [\![1,N]\!]} \subseteq \{0,1\}$ and $\{R_2[x,y]\}_{x \in [\![1,N]\!], y \in [\![1,N]\!]} \subseteq \{0,1\}$, where N is the mesh's resolution, the following comparator is implemented and used to sort the list of regions:

$$\text{Let } \text{diff}(R_1, R_2) = \sum_{x=1}^{N}\sum_{y=1}^{N} R_1[x,y] - R_2[x,y] \implies \begin{cases} \text{diff}(R_1, R_2) < 0 \implies R_1 \text{ under } R_2, \\ \text{diff}(R_1, R_2) > 0 \implies R_2 \text{ under } R_1. \end{cases}$$

3.3 COMPUTATION OF SPECIFIC MECHANISMS' PRIVACY REGIONS

We aim to compute privacy regions of mechanisms through their trade-off functions. Later, we will compare the actual privacy guarantees they offer against their DP guarantees shown in 2.4. We will also involve them in utility-privacy trade-offs.

3.3.1 LAPLACE MECHANISM

First, we compute the trade-off function of the Laplace mechanism, defined in 2.4.2.

PROPOSITION 38: LAPLACE MECHANISM TRADE-OFF FUNCTION
Let $f : \mathcal{X}^n \to \mathbb{R}^k$ be a function with finite ℓ_1-sensitivity, and $\varepsilon > 0$. Then the Laplace mechanism as described in 2.4.2 is $L-$DP with

$$L(\alpha) = F_{\text{Laplace}(0,1)}\left(F^{-1}_{\text{Laplace}(0,1)}(1-\alpha) - \varepsilon\right).$$

PROOF: Let Δ be the ℓ_1-sensitivity of f, i.e

$$\Delta = \Delta_{f,1}.$$

By definition,

$$\begin{aligned}
L(\alpha) &= \inf_{\substack{X_0,X_1\in\mathcal{X}^n \\ X_0,X_1 \text{ neighbouring}}} T\left(\prod_{i=1}^k \text{Laplace}\left(f(X_0)_i, \frac{\Delta}{\varepsilon}\right), \prod_{i=1}^k \text{Laplace}\left(f(X_1)_i, \frac{\Delta}{\varepsilon}\right)\right)(\alpha) \\
&= \inf_{\substack{X_0,X_1\in\mathcal{X}^n \\ X_0,X_1 \text{ neighbouring}}} T\left(\prod_{i=1}^k \text{Laplace}\left(f(X_0)_i - f(X_1)_i, \frac{\Delta}{\varepsilon}\right), \prod_{i=1}^k \text{Laplace}\left(0, \frac{\Delta}{\varepsilon}\right)\right)(\alpha) \\
&= \inf_{\substack{X_0,X_1\in\mathcal{X}^n \\ X_0,X_1 \text{ neighbouring}}} T\left(\prod_{i=1}^k \text{Laplace}\left(f(X_0)_i - f(X_1)_i, \frac{\Delta}{\varepsilon}\right), \prod_{i=1}^k \text{Laplace}\left(0, \frac{\Delta}{\varepsilon}\right)\right)(\alpha) \\
&= T\left(\text{Laplace}\left(\Delta, \frac{\Delta}{\varepsilon}\right), \text{Laplace}\left(0, \frac{\Delta}{\varepsilon}\right)\right)(\alpha) \text{ using } k\text{Laplace}(\mu, b) = \text{Laplace}(k\mu, |k|b), \\
&= T\left(\frac{\Delta}{\varepsilon}\text{Laplace}(\varepsilon, 1), \frac{\Delta}{\varepsilon}\text{Laplace}(0,1)\right)(\alpha) \\
&= T\left(\text{Laplace}(\varepsilon,1), \text{Laplace}(0,1)\right)(\alpha) \\
&= F_{\text{Laplace}(0,1)}\left(F^{-1}_{\text{Laplace}(0,1)}(1-\alpha) - \varepsilon\right).
\end{aligned}$$

□

Observe that, with this parametrisation, the Laplace mechanism's privacy guarantee does not depend on $\Delta_{f,1}$, f's sensitivity. $\Delta_{f,1}$ will however play a role later when we examine utility-privacy trade-offs with the Laplace mechanism.

3.3.2 GAUSSIAN MECHANISM

As for the Laplace mechanism, we start by computing the trade-off function of the mechanism as prescribed by 2.4.3.

PROPOSITION 39: GAUSSIAN MECHANISM TRADE-OFF FUNCTION
Let $f : \mathcal{X}^n \to \mathbb{R}^k$ be a function with finite ℓ_2-sensitivity, $\varepsilon > 0$ and $\delta \in\ ]0, 1]$. Then the Gaussian mechanism as described in 2.4.3 is μ-GDP with $\mu = \frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}$, i.e it is $G_{\frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}}$-DP with

$$G_{\frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}}(\alpha) = \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}\right).$$

This proof resembles the previous one, for completeness it can be found in appendix B.2.1.

3.3.3 RANDOMIZED RESPONSE

The privacy region of randomized response is essentially computed in [9].

PROPOSITION 40: RANDOMIZED RESPONSE EXACT PRIVACY REGION
Let $\varepsilon \geq 0$ and let $\mathcal{X}$ be a finite set. Then randomized response on $\mathcal{X}$ with parameter ε is $(\varepsilon, 0)-$DP and η-TV with

$$\eta = \frac{e^\varepsilon - 1}{e^\varepsilon + |\mathcal{X}| - 1}.$$

3.3.4 EXPONENTIAL MECHANISM

Computing a closed form for the exact privacy region of an arbitrary exponential mechanism is intractable. Nevertheless, the following generic lemma is useful to compute privacy regions of specific instances of the exponential mechanisms. Its proof is deferred to appendix B.2.1.

LEMMA 41: OPTIMAL TEST REGION FOR THE EXPONENTIAL MECHANISM
Let M be the exponential mechanism on $\mathcal{X}^n$ with score function $s : \mathcal{X}^n \times \mathcal{C} \to \mathbb{R}$ and DP parameter $\varepsilon > 0$. Let X_0, X_1 be neighbouring datasets in $\mathcal{X}^n$. Fix a maximum false positive error rate $\alpha \in\ ]0, 1[$, i.e, for a test region $T \subseteq \mathcal{C}$, require that

$$\mathrm{FP}\left(X_0, X_1, M, T\right) = \mathbb{P}\left(M(X_0) \in T\right) \leq \alpha.$$

Then, there exists $t_\alpha(X_0, X_1) \in \mathbb{R}$ such that the optimal T, minimizing the false-negative probability $\mathrm{FN}\left(X_0, X_1, M, T\right) = \mathbb{P}\left(M(X_1) \in T^C\right)$, is

$$T = \left\{c \in \mathcal{C} \mid s(X_1, c) - s(X_0, c) \geq t_\alpha(X_0, X_1)\right\}.$$

The $(\varepsilon, 0)-$DP guarantee can also be sufficient for some applications, as we will see later.

3.4 VISUALISING PRIVACY REGIONS

3.4.1 PRIVACY REGIONS WINDOW'S FEATURES

The initial state of the privacy regions window is depicted in 3.3.

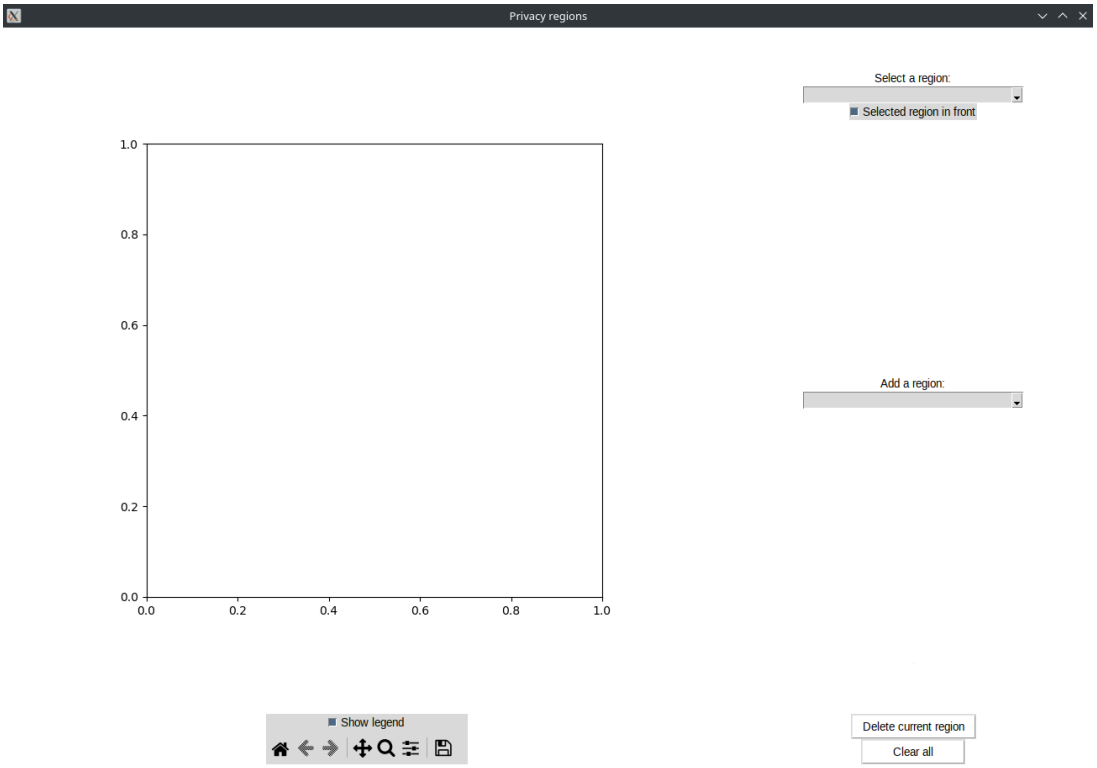

FIGURE 3.3
Initial state of the privacy regions window

The user is expected to add a region using the "Add a region" combobox, on the right-hand pane. Figure 3.4 shows the default DP-TV region ($\varepsilon = 0.6, \delta = 0.15, \eta = 0.25$). Then, the influence of each of the parameters $\varepsilon, \delta, \eta$ can be visualised by playing with the sliders in the bottom-right corner. The user can also add more regions using the "Add a region" combobox. In figure 3.5, a 1.5-GDP region.

To modify a region's parameters, one can select it through the "Select a region" combobox in the upper right corner. Once selected, a region can also be deleted by clicking the "Delete current region" button. The checkbox "Selected region in front" toggles whether or not the selected region appears in front of all the other regions in the figure.

It is also possible to intersect regions. Select "Intersect regions" in the "Add a region" combobox. A window such as 3.6 opens, prompting to choose currently drawn regions to be intersected along with a title for the intersection, defaulting to "Intersection". In this instance, intersecting an ($\varepsilon = 0.6, \delta = 0.15$)-DP region with $\eta = 0.25$-TV with a 0.8-GDP region yields figure 3.7.

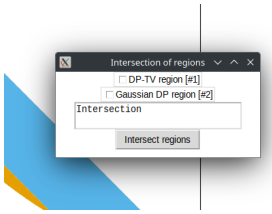

FIGURE 3.6
Intersection selection window

3.4.2 VISUALISING VARIATIONS OF DIFFERENTIAL PRIVACY

Throughout this thesis, we have shown some figures depicting the privacy regions for GDP and DP-TV, such as 2.1 and 2.5. These figures were obtained using the tool, as the navigation bar under

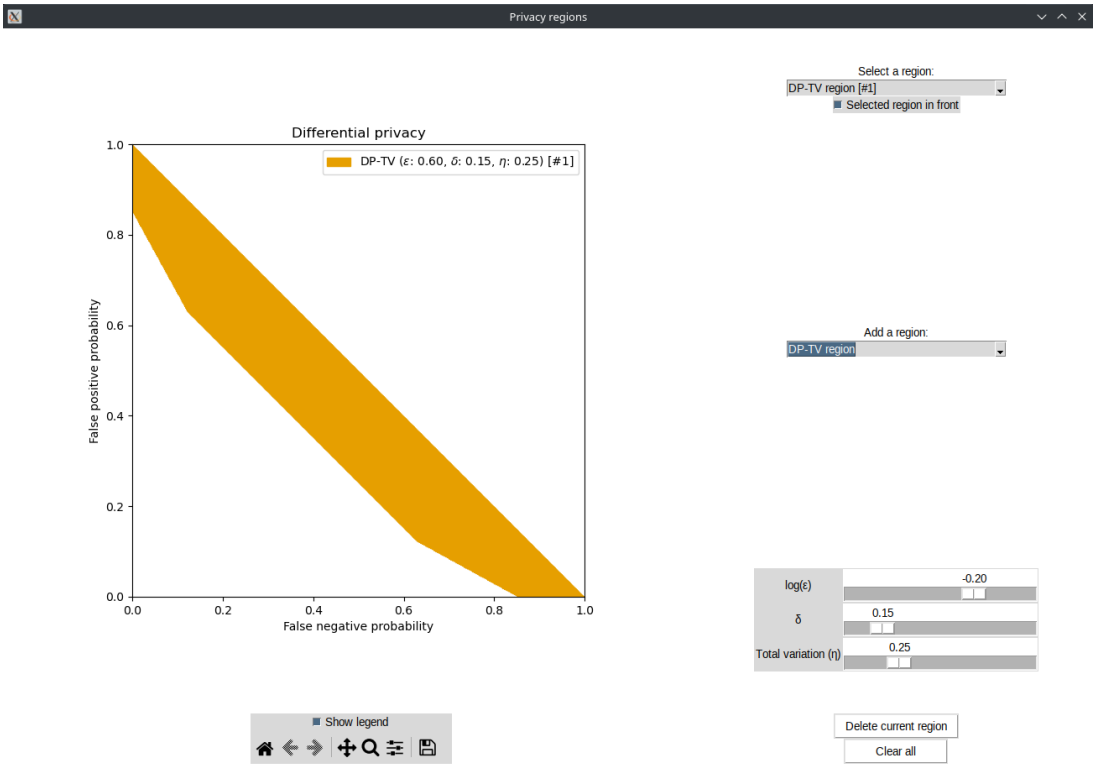

FIGURE 3.4
Added region to privacy regions window

the privacy plot has a save button.

3.4.3 REGIONS FOR SPECIFIC MECHANISMS

The privacy regions computed in 3.3 are readily available as regions that can be added via the "Add a region" combobox. As an example, the Laplace mechanism's exact privacy region is compared to its corresponding DP-TV guarantee in 3.8.

Such comparisons of mechanisms' exact regions against their DP guarantees and variations thereof is an excellent use case for the tool, as the user can gain intuition on how each parameter influences said privacy regions.

3.4.4 VISUAL ASSESSMENT OF COMPOSITION THEOREMS

Another important use case of the privacy-regions window is the visualisation of composition theorems. As seen in sections 2.1.4 and 2.5, these theorems can be hard to parse. Visualisation helps us appreciate how a composition result qualitatively and quantitatively differs from another, hence one can readily plot these composition regions using the tool.

Figure 3.9 compares the regions for the composition of 5 $(0.6, 0.05)-$DP mechanisms with 0.15 total variation, with the basic DP (2.3.1, Theorem 10), exact DP (2.3.2, Theorem 11) and exact DP-TV theorems (2.5.1, Theorem 29).

Also, figure 3.10 showcases the convergence of the composition of $(\varepsilon, 0)-$DP mechanisms to a $\mu-$GDP privacy guarantee, as obtained through Theorem 37 of 2.5.2. The authors of [10] make this convergence more explicit through the following corollary.

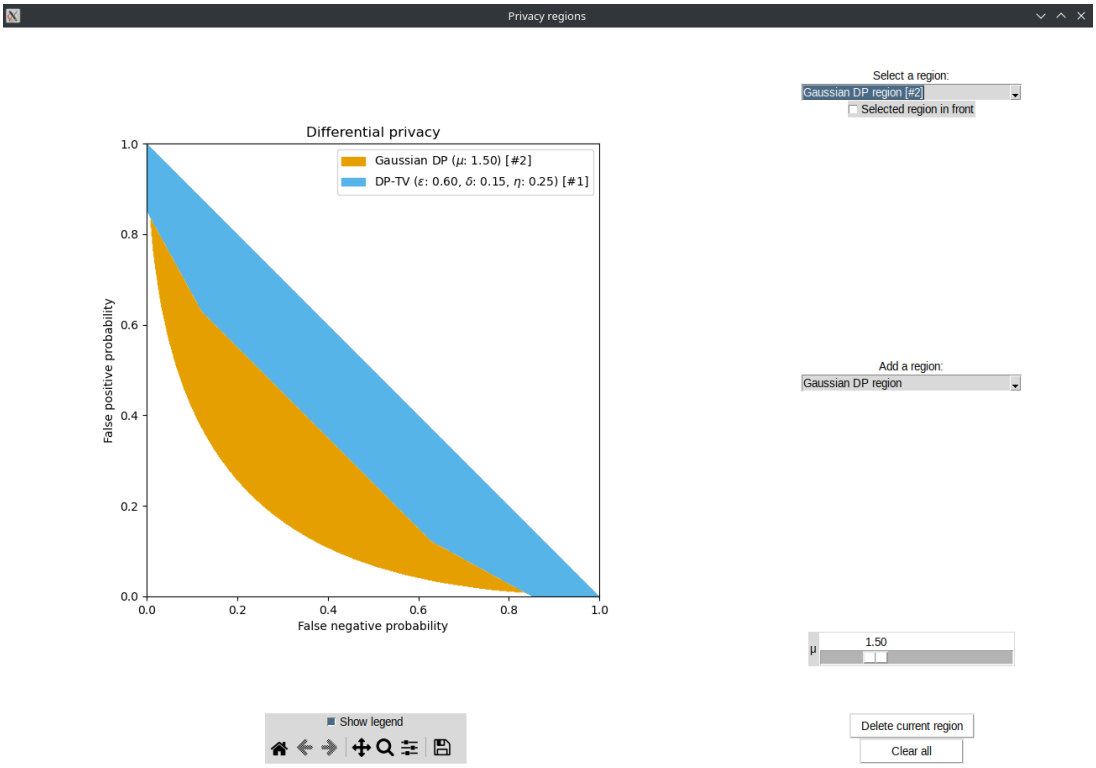

FIGURE 3.5
Two privacy regions

COROLLARY 42: $(\varepsilon, \delta)-$DP CONVERGES TO $\mu-$GDP TIMES UNIFORM
Let $\{(\varepsilon_{n,i})_{i=1}^n\}_{n\in\mathbb{N}^*}$ and $\{(\delta_{n,i})_{i=1}^n\}_{n\in\mathbb{N}^*}$ be triangular arrays such that

- $\sum_{i=1}^n \varepsilon_{n,i}^2 \to \mu^2$ and $\max_{i\in[\![1,n]\!]} \varepsilon_{n,i} \to 0$,
- $\sum_{i=1}^n \delta_{n,i}^2 \to \delta$ and $\max_{i\in[\![1,n]\!]} \delta \to 0$.

Then

$$\bigotimes_{i=1}^n f_{\varepsilon_{n,i},\delta_{n,i}} \overset{n\to\infty}{\to} G_\mu \otimes T\left(\mathrm{Unif}(0,1), \mathrm{Unif}(1-e^{-\delta}, 2-e^{-\delta})\right)$$

uniformly over $[0,1]$.

Recall the definition of $f_{\varepsilon,\delta}$ from corollary 32 in section 2.5.2.

3.5 VISUALISING PRIVACY-UTILITY TRADE-OFFS

3.5.1 TRADE-OFFS' WINDOW'S FEATURES

After choosing a combination of a query, mechanism and main privacy parameter through the main menu, a privacy-utility trade-off window opens as depicted in figure 3.2. The left-hand figure plots the utility function against the chosen main privacy parameter, here δ, along with a dotted line marking that parameter's current value. The right-hand figure is the privacy region of the mechanism parametrised by the current value of the privacy parameters, in this example ε and δ. Also, the bottom-left corner features sliders that change the value of the privacy parameters as well as other relevant parameters, such as the

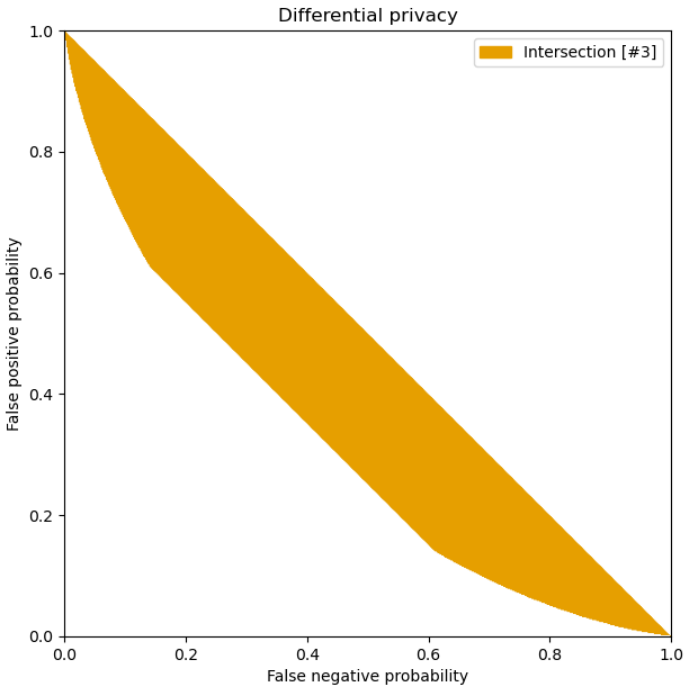

FIGURE 3.7
Intersection region

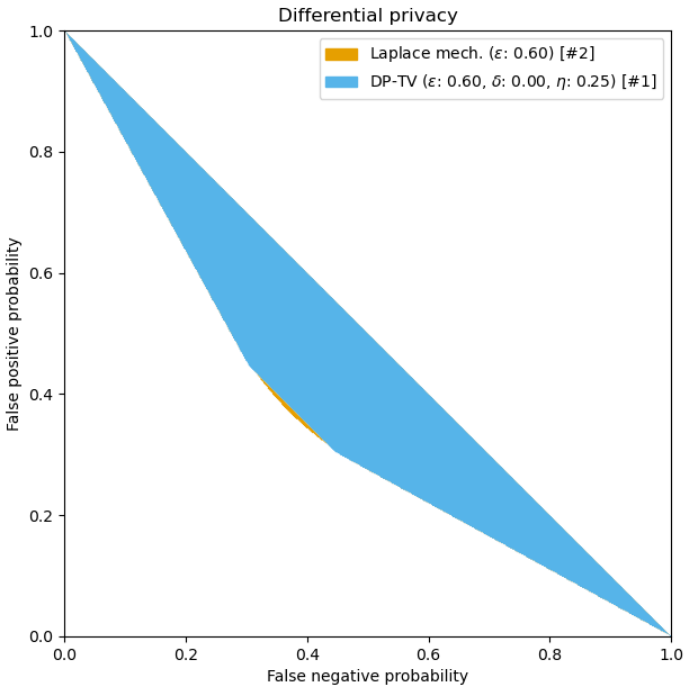

FIGURE 3.8
Comparing the Laplace mechanism's privacy region to its DP-TV guarantee

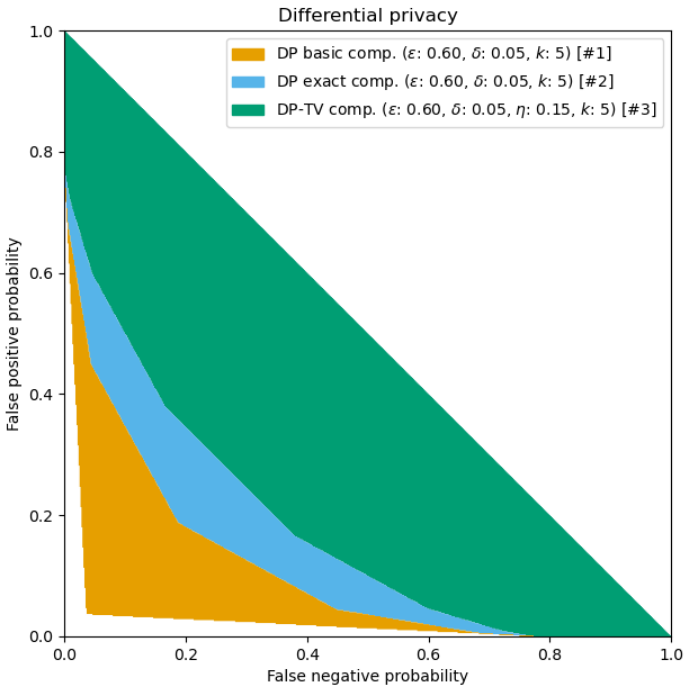

FIGURE 3.9
Comparison of basic, exact DP and exact DP-TV regions

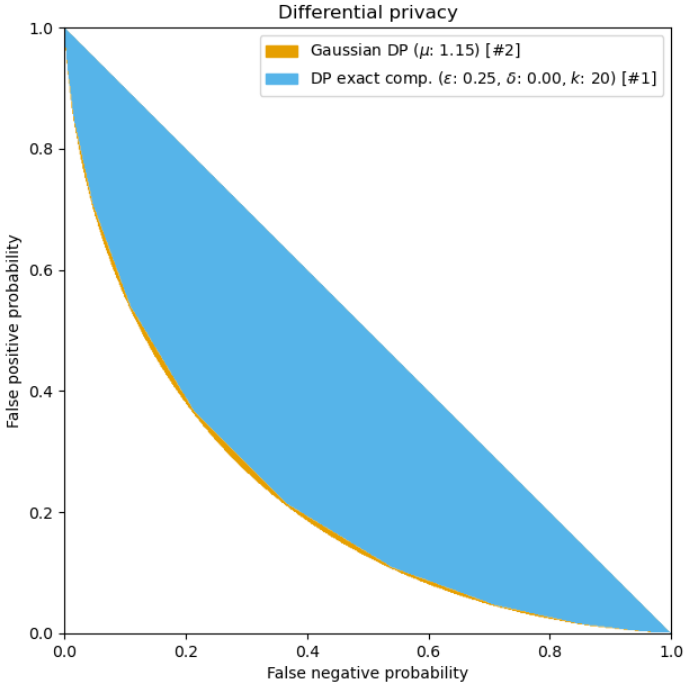

FIGURE 3.10
Convergence of the region for the composition of $(\varepsilon, 0)-$DP mechanisms to a $\mu-$GDP guarantee

size of the dataset. Any change in the value of a slider triggers the redrawing of the utility, and changing the value of a privacy parameter's slider redraws the privacy region.

In what follows, we present the implemented privacy-utility trade-off combinations of query, mechanism and utility.

3.5.2 HISTOGRAM WITH LAPLACE MECHANISM

First, we define the **histogram** query and compute its sensitivity.

DEFINITION 43: HISTOGRAM, PROBABILITY SIMPLEX
Let $\mathcal{X}$ be a set and $n \in \mathbb{N}^*$. Partition $\mathcal{X}$ into k disjoint bins

$$\mathcal{X} = \bigsqcup_{i=1}^{k} \mathcal{B}_i.$$

Denote the $k-$**dimensional probability simplex** by $\mathcal{S}_k = \left\{ p \in \mathbb{R}_+^k \mid \sum_{i=1}^k p_i = 1 \right\}$.
Thus the **histogram** query is the function

$$\begin{aligned} \text{Hist}_{\{\mathcal{B}_i\}} : \mathcal{X}^n &\to \mathcal{S}_k \\ X &\mapsto \frac{1}{n} \begin{bmatrix} \sum_{i=1}^n \mathbb{1}\{X_i \in \mathcal{B}_1\} \\ \sum_{i=1}^n \mathbb{1}\{X_i \in \mathcal{B}_2\} \\ \vdots \\ \sum_{i=1}^n \mathbb{1}\{X_i \in \mathcal{B}_k\} \end{bmatrix}. \end{aligned}$$

LEMMA 44: SENSITIVITY OF THE HISTOGRAM QUERY
$\text{Hist}_{\{\mathcal{B}_i\}}$ has sensitivity $\Delta_1 = 2$.

PROOF: The histogram query counts the number of elements of X falling into each of the k bins. By changing one element in X_0 and obtaining X_1, one can at most remove one element from a bin and add it to another, changing the difference $||\text{Hist}_{\{\mathcal{B}_i\}}(X_0) - \text{Hist}_{\{\mathcal{B}_i\}}(X_1)||_1$ by 2. □

Thus we can now define the **differentialy-private histogram** query, or DP histogram, which uses the previously defined Laplace mechanism (2.4) with the computed sensitivity.

DEFINITION 45: DP HISTOGRAM

Let $\mathcal{X}$ be a set, $n \in \mathbb{N}^*$ and $\varepsilon > 0$. Let $\mathcal{X} = \bigsqcup_{i=1}^{k} \mathcal{B}_i$. We define the **DP histogram** $\text{DPHist}_{\varepsilon,\{\mathcal{B}_i\}}$ as the Laplace mechanism applied on $f = \text{Hist}_{\{\mathcal{B}_i\}}$. Explicitly,

$$\begin{aligned} \text{DPHist}_{\varepsilon,\{\mathcal{B}_i\}} : \mathcal{X}^n &\to \mathbb{R}^k \\ X &\mapsto \text{Hist}_{\{\mathcal{B}_i\}}(X) + (Y_1, \dots, Y_k) \end{aligned}$$

where

$$\{Y_i\}_{i=1}^{k} \overset{\text{i.i.d}}{\sim} \text{Laplace}\left(0, \frac{2}{\varepsilon}\right).$$

Lastly, we set a meaningful utility for this problem. It is tempting to use the **mean absolute error**, since the added noise is Laplace-distributed. However, we argue that this is not necessarily the best choice in this instance: it linearly punishes deviations from the true mean, allowing for potentially large errors in the distribution estimation for one particular bin. But since the true mean, i.e the non private histogram, is a vector in $\mathcal{S}_k$, its coefficients are already fairly small as they are in $[0, 1]$. Thus the **mean squared error** (MSE) is preferable, as it punishes large deviations more severely, avoiding estimates which are too far off for a particular bin. We compute a closed form for the MSE.

PROPOSITION 46: MSE OF DPHIST

In the same setup as the previous definition,

$$\text{MSE}\left(\text{Hist}_{\{\mathcal{B}_i\}}(X), \text{DPHist}_{\varepsilon,\{\mathcal{B}_i\}}(X)\right) := \mathbb{E}\left(||\text{Hist}_{\{\mathcal{B}_i\}}(X) - \text{DPHist}_{\varepsilon,\{\mathcal{B}_i\}}(X)||_2^2\right) = \frac{8k}{\varepsilon^2}.$$

PROOF:

$$\begin{aligned} \mathbb{E}\left(||\text{Hist}_{\{\mathcal{B}_i\}}(X) - \text{DPHist}_{\varepsilon,\{\mathcal{B}_i\}}(X)||_2^2(X)\right) &= \mathbb{E}\left(||\text{Hist}_{\{\mathcal{B}_i\}}(X) - \text{Hist}_{\{\mathcal{B}_i\}}(X) - (Y_i)_{i=1}^{k}||_2^2\right) \\ &= \mathbb{E}\left(||(Y_i)_{i=1}^{k}||_2^2\right) \\ &= \sum_{i=1}^{k} \mathbb{E}\left(Y_i^2\right) \\ &= k\text{Var}\left(\text{Laplace}\left(0, \frac{2}{\varepsilon}\right)\right) \text{ (identical distribution)} \\ &= \frac{8k}{\varepsilon^2}. \end{aligned}$$

□

Since the MSE $\in \Theta(\frac{1}{\varepsilon^2})$, as can be observed in figure 3.11, it decreases relatively sharply: linearly in a doubly-logarithmic plot. Nonetheless, to make the error low in absolute terms, one has to give up a lot of privacy.

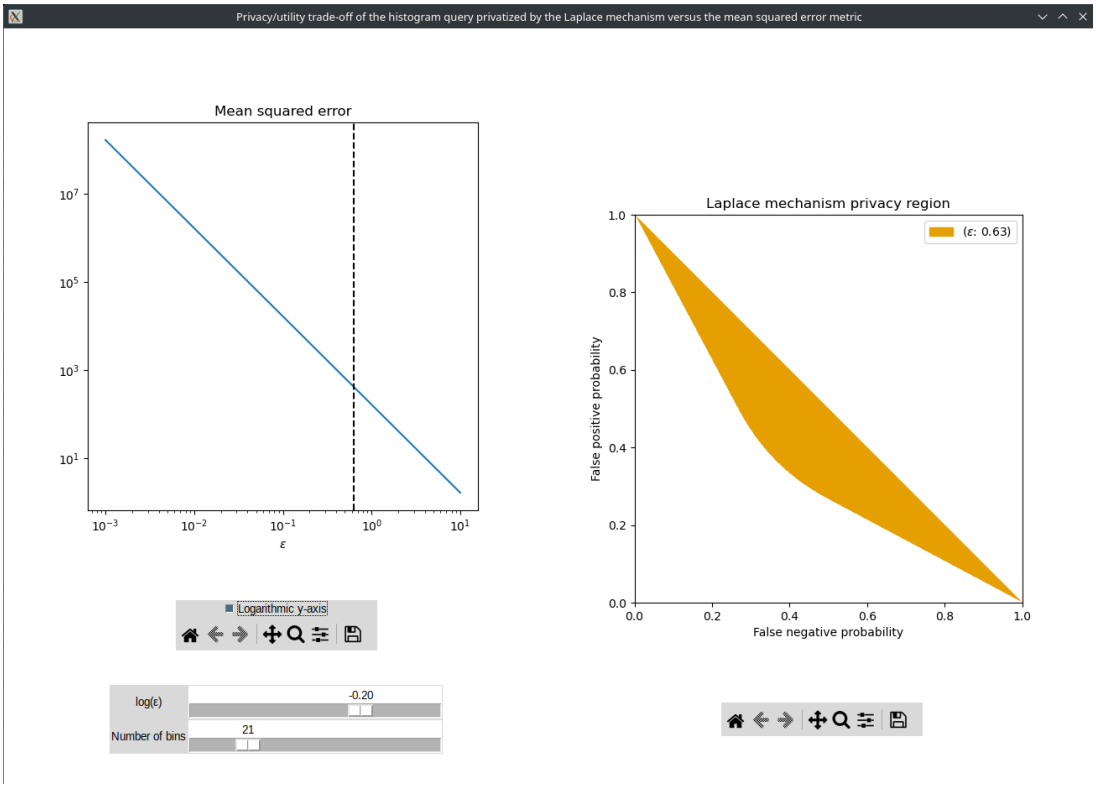

FIGURE 3.11
Histogram privacy-utility window

3.5.3 MEAN WITH GAUSSIAN MECHANISM

We give another example of a mechanism with an MSE utility, the differentially private mean as an application of the Gaussian mechanism (2.4).

DEFINITION 47: DIAMETER OF A SET
Let $\mathcal{X} \subseteq \mathbb{R}^k$. We define the **diameter** of $\mathcal{X}$ to be

$$d := d(\mathcal{X}) = \sup_{a,b \in \mathcal{X}} ||a - b||_2.$$

LEMMA 48: SENSITIVITY OF THE MEAN QUERY
Let $\mathcal{X} \subset \mathbb{R}^k$ of diameter d and $n \in \mathbb{N}^*$. We define the **mean** query

$$\text{Mean} : \mathcal{X}^n \to \mathbb{R}^k$$
$$X \mapsto \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Mean has sensitivity $\Delta_2 = \frac{d}{n}$.

PROOF: Let $X_0, X_1 \in \mathcal{X}^n$ be neighbouring datasets. Assume w.l.o.g that $X_{0,1} \neq X_{1,1}$.

$$\begin{aligned} ||\text{Mean}(X_0) - \text{Mean}(X_1)||_2 &= \frac{1}{n}\sqrt{\sum_{i=1}^{n}(X_{0,i} - X_{1,i})^2} \\ &= \frac{1}{n}|X_{0,1} - X_{1,1}| \text{ (all indices equal except at index 1),} \end{aligned}$$

thus $\sup_{X_0, X_1 \text{ neighbouring}} ||\text{Mean}(X_0) - \text{Mean}(X_1)||_2 = \frac{1}{n}\sup_{X_{0,1} \neq X_{1,1}} |X_{0,1} - X_{1,1}| = \frac{d}{n}$. □

DEFINITION 49: DP MEAN

Let $\mathcal{X} \subset \mathbb{R}^k$ of diameter d, $n \in \mathbb{N}^*$, $\varepsilon > 0$ and $\delta \in \ ]0, 1]$. We define the **DP mean** $\text{DPMean}_{\varepsilon,\delta,d}$ as the Gaussian mechanism applied on the mean. Explicitly,

$$\begin{aligned} \text{DPMean}_{\varepsilon,\delta} : \mathcal{X}^n &\to \mathbb{R}^k \\ X &\mapsto \frac{1}{n}\sum_{i=1}^{n} X_i + (Y_1, \ldots, Y_k) \end{aligned}$$

where

$$\{Y_i\}_{i=1}^{k} \overset{\text{i.i.d}}{\sim} \mathcal{N}\left(0, 2\log\left(\frac{5}{4\delta}\right)\frac{d^2}{n^2\varepsilon^2}\right).$$

Lastly, we compute the MSE. We defer the proof to appendix B.2.2 as it is similar to the Laplace case.

PROPOSITION 50: MSE OF DPMEAN

In the same setup as the previous definition,

$$\text{MSE}(\text{Mean}, \text{DPMean}_{\varepsilon,\delta}) = 2k\log\left(\frac{5}{4\delta}\right)\frac{d^2}{n^2\varepsilon^2}.$$

3.5.4 MEDIAN WITH EXPONENTIAL MECHANISM

Here, we will look at the differentially private computation of the **median** of a dataset along with an interesting utility.

PROPOSITION 51: DIFFERENTIALLY PRIVATE MEDIAN
Let $\mathcal{X} = [\![1, m]\!]$ and $\varepsilon > 0$. Define the score function

$$q : \mathcal{X}^n \times \mathcal{X} \to \mathcal{X}$$
$$(X, c) \mapsto q(X, c) = -\left|\sum_{i=1}^{n} \text{sign}\,(c - X_i)\right|.$$

Then the exponential mechanism M instantiated with score function q and DP parameter ε has the following utility guarantee. Let $\text{position}(c, X)$ denote the position of c in the sorted order of X. Then

$$\mathbb{P}\left(\left|\text{position}(M(X), X) - \frac{n}{2}\right| \geq t\right) \leq m \exp\left(-\frac{\varepsilon}{4}t\right).$$

Before proving this, we need to compute a bound on the sensitivity of q.

LEMMA 52:
q has sensitivity $\Delta \leq 2$.

A proof of this lemma can be found in appendix B.2.2. Now for the proof of the aforementioned proposition.

PROOF: For simplicity, assume that the median m of X is balanced, in the sense that there exists $m \in X$ such that $q(X, m) = 0$. The proof is similar if it is not the case.
It suffices to apply the exponential mechanism's utility guarantee proposition on q to get:

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \text{sign}\,(M(X) - X_i)\right| \geq \frac{2\Delta}{\varepsilon}(\log(m) + \tau)\right) \leq \exp(-\tau).$$

Indeed, the optimal value for q is 0. Also recall that $\Delta = 2$. Then, to conclude, it suffices to observe the following. $\forall c \in [\![1, m]\!]$,

$$\left|\text{position}(c, X) - \frac{n}{2}\right| = \left|\sum_{i=1}^{n} \text{sign}\,(c - X_i)\right|.$$

Indeed, both sides of this equation are counting the number of the positions that c is off of the true median. Hence,

$$\mathbb{P}\left(\left|\text{position}(M(X), X) - \frac{n}{2}\right| \geq t\right) = \mathbb{P}\left(\left|\sum_{i=1}^{n} \text{sign}\,(M(X) - X_i)\right| \geq t\right).$$

astly, to find the conjectured bound, set $\tau = \frac{\varepsilon}{4}t - \log(m)$ in the first inequality. □

The exponential decay on the probability's bound implies that, for t large enough, the bound drops swiftly to zero with acceptable DP guarantees for many applications. Observe that the bound does not depend directly on the dataset size n, and only indirectly through t. Thus this bound is especially good for n large, e.g 10^5 or more, as the bound is essentially 0 for t several orders of magnitude smaller. Figure 3.12 illustrates this situation.

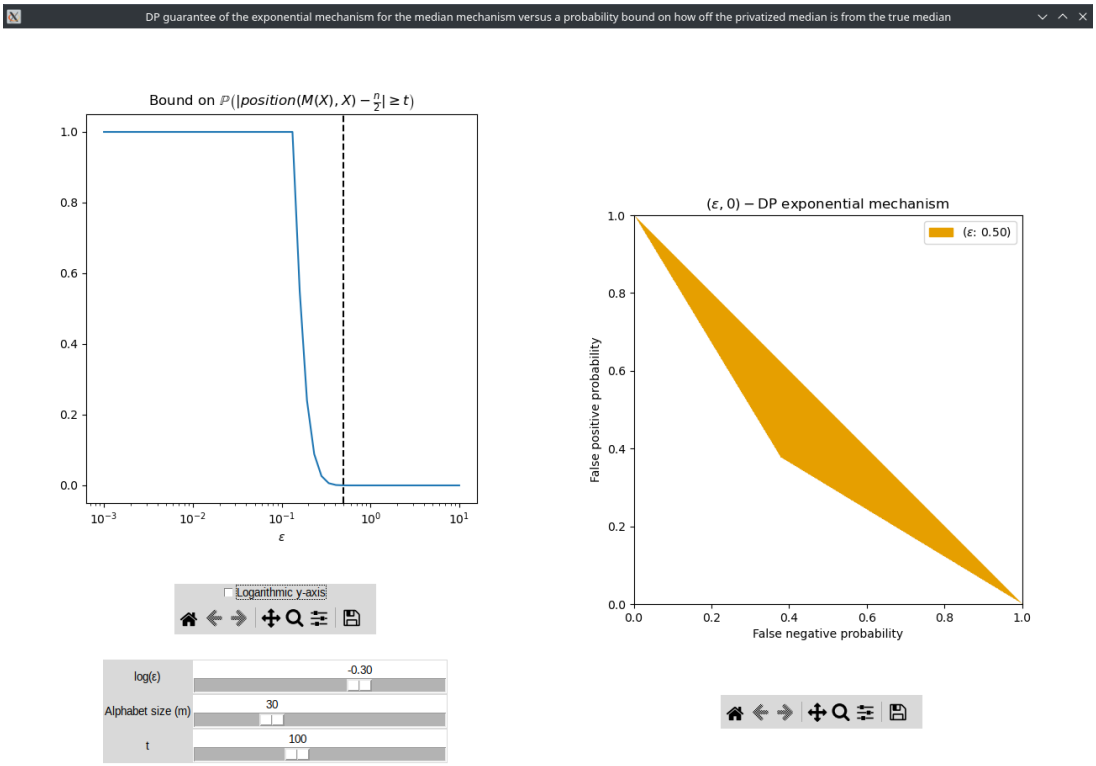

FIGURE 3.12
Median privacy-utility window

3.5.5 RANDOMIZED RESPONSE

As mentioned in section 3.3.3, the privacy region for randomized response is a DP-TV region. As an (inverse) utility, we put forward the probability of making a uniformly random choice as defined in 2.4.4,

$$1 - p_\varepsilon = 1 - \frac{e^\varepsilon - 1}{e^\varepsilon + |\mathcal{X}| - 1} = \frac{|\mathcal{X}|}{e^\varepsilon + |\mathcal{X}| - 1}.$$

Figure 3.13 displays this privacy-utility trade-off. As randomized response's exact privacy region corresponds exactly to DP-TV, an absolutely low p_ε comes at a major price in privacy.

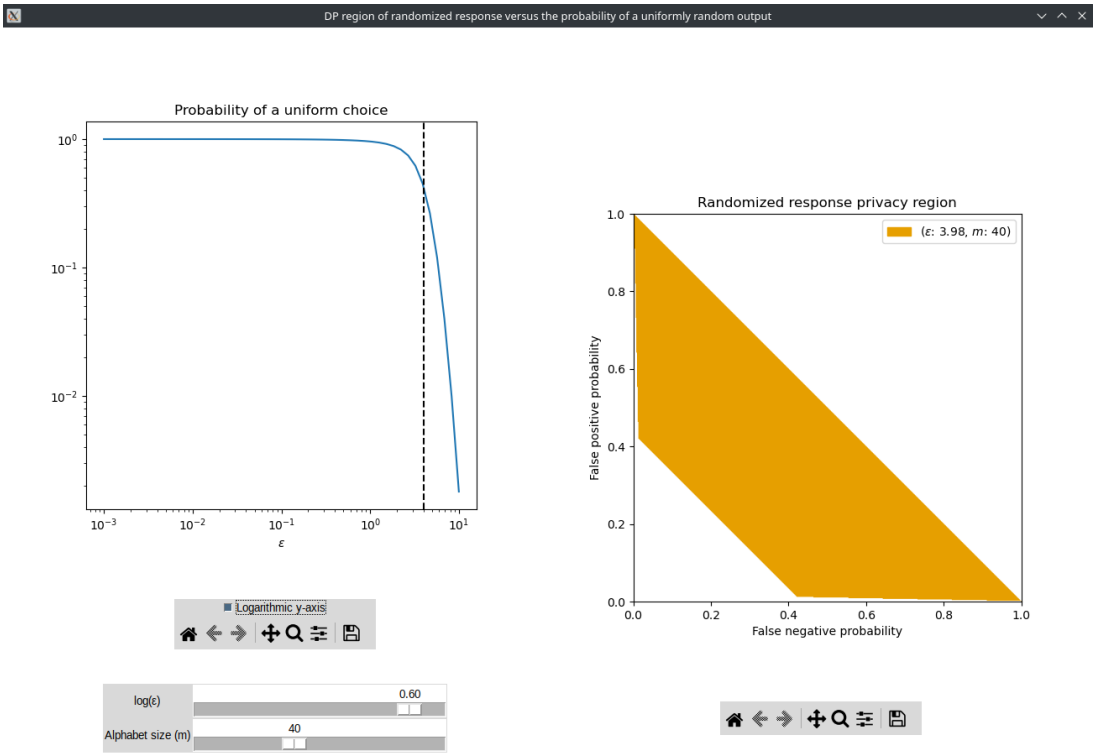

FIGURE 3.13
Randomized response privacy-utility window

CHAPTER 4

NEW RESEARCH QUESTIONS ARISING FROM VISUALISATION

While working on the presented visualisation tool, thinking about and reading the associated literature, we have explored new research directions in DP. In some instances, visualisation helped us in this endeavour. This chapter is a report on some of these new ideas, specifically the ones related to the composition of DP mechanisms.

4.1 COMPOSITION OF MECHANISMS WITH MULTIPLE DP CONSTRAINTS

4.1.1 RECASTING THE COMPOSITION THEOREM FOR DIFFERENTIAL PRIVACY WITH TOTAL VARIATION

We revisit theorem 29 in section 2.5.1, i.e the composition theorem for DP-TV mechanisms. Since

$$d_{\mathrm{TV}}(M) \leq \eta \iff M \text{ is } (0,\eta) - \mathrm{DP},$$

we can state the theorem as the composition of mechanisms for which **two** DP constraints are known to hold: each M_i is both $(\varepsilon_1, \delta_1) - \mathrm{DP}$ and $(0, \delta_2) - \mathrm{DP}$.

THEOREM 53: RESTATING THE COMPOSITION THEOREM FOR DP-TV
Let $M_1, \ldots, M_k$ be $(\varepsilon_1, \delta_1)-$DP and $(0, \delta_2)-$DP mechanisms with independent internal randomness where $\varepsilon_1 \geq 0$, $\delta_1 \in [0,1]$ and $\delta_2 \in \left[\delta_1, \delta_1 + (1-\delta_1)\frac{e^{\varepsilon_1}-1}{e^{\varepsilon_1}+1}\right]$. Then their adaptive composition

$$M = (M_1, \ldots, M_k) \text{ is } \left(j\varepsilon_1, 1 - (1-\delta_1)^k(1-\delta^{(j)})\right) - \mathrm{DP} \text{ and } \eta - \mathrm{TV}$$

for all $j \in [\![0, k]\!]$, where

$$\delta^{(j)} = \sum_{i=0}^{k-j-1} \binom{k}{i} \sum_{l=0}^{\lceil \frac{k-j-i}{2} \rceil - 1} \binom{k-i}{l} \left(\frac{\mu-1}{e^{\varepsilon_1}+1}\right)^{k-i} \mu^i \left(e^{(k-l-i)\varepsilon_1} - e^{(l+j)\varepsilon_1}\right),$$

$$\mu = 1 - \frac{\delta_1 - \delta_2}{\delta_1 - 1} \cdot \frac{e^{\varepsilon_1}-1}{e^{\varepsilon_1}+1}, \text{ and}$$

$$\eta = 1 - (1-\delta_1)^k(1-\delta^{(0)}).$$

As a natural extension of this result, we seek the DP guarantees on the composition of k mechanisms M_i such that

$$\text{each } M_i \text{ is both } (\varepsilon_1, \delta_1)-\text{DP and } (\varepsilon_2, \delta_2)-\text{DP with } \varepsilon_2 \geq 0.$$

4.1.2 ATTEMPTS AT FINDING THE COMPOSITION REGION

For notational ease, we first define the **privacy region of a trade-off function**.

DEFINITION 54: PRIVACY REGION OF A TRADE-OFF FUNCTION
Let f be a trade-off function. Then we define the **privacy region of** f to be

$$\mathcal{R}(f) = \left\{ (\text{FN}, \text{FP}) \in [0,1]^2 \mid \text{FN} \geq f(\text{FP}) \right\}.$$

To gain some intuition on the problem, we set $k = 2$, i.e we assume the number of mechanisms to be composed is 2. Let M_a, M_b be two mechanisms which are both $(\varepsilon_1, \delta_1)-$DP and $(\varepsilon_2, \delta_2)-$DP. An initially compelling hypothesis for the privacy region of the composition was the intersection of the 2-compositions of all the pairs of privacy guarantees, as follows:

$$\mathcal{R}((M_a, M_b)) \overset{?}{\subseteq} \mathcal{R}(f_{\varepsilon_1,\delta_1} \otimes f_{\varepsilon_1,\delta_1}) \cap \mathcal{R}(f_{\varepsilon_1,\delta_1} \otimes f_{\varepsilon_2,\delta_2}) \cap \mathcal{R}(f_{\varepsilon_2,\delta_2} \otimes f_{\varepsilon_2,\delta_2}),$$

and we have postulated the existence a pair of mechanisms M_a, M_b that have the required properties and attain this bound.

This is a convenient bound as $\mathcal{R}(f_{\varepsilon_1,\delta_1} \otimes f_{\varepsilon_1,\delta_1})$ is known from theorem 11 in section 2.3.2 and $\mathcal{R}(f_{\varepsilon_1,\delta_1} \otimes f_{\varepsilon_2,\delta_2})$ can be well approximated using theorem 12. To check the tightness hypothesis, we fall back to the known case $\varepsilon_2 = 0$ where the optimal region is known from [9]. In that case, the hypothesised region simplifies to

$$\mathcal{R}((M_a, M_b)) \overset{?}{\subseteq} \mathcal{R}(f_{\varepsilon_1,\delta_1} \otimes f_{\varepsilon_1,\delta_1}) \cap \mathcal{R}(f_{\varepsilon_1,\delta_1} \otimes f_{0,\delta_2}) \cap \mathcal{R}(f_{0,\delta_2} \otimes f_{0,\delta_2}),$$

where

- $\mathcal{R}(f_{\varepsilon_1,\delta_1} \otimes f_{\varepsilon_1,\delta_1})$ is again computed exactly using theorem 11,
- $\mathcal{R}(f_{\varepsilon_1,\delta_1} \otimes f_{0,\delta_2}) = \mathcal{R}(\varepsilon_1, 1-(1-\delta_1)(1-\delta_2))$. Indeed, interpreting the δ in the definition of (ε, δ) differential privacy as the probability of leaking data, as explained in [2],
- $\mathcal{R}(f_{0,\delta_2} \otimes f_{0,\delta_2}) = \mathcal{R}\left(0, 1-(1-\delta_2)^2\right)$.

To show the latter two equalities, we apply the following proposition.

PROPOSITION 55:
The composition of an $(\varepsilon, \delta_1)-$DP mechanism with a $(0, \delta_2)-$DP mechanism with independent internal randomness is $(\varepsilon, 1-(1-\delta_1)(1-\delta_2))-$DP.

PROOF: Let M_1 and M_2 refer to the two aforementioned mechanisms respectively. In [6], the interpretation of the δ parameter as a *pure data leakage probability* is presented:

$$M \text{ is } (\varepsilon, \delta) - \text{DP} \implies \mathbb{P}(M \text{ discloses data}) \leq \delta.$$

Thus, using the independence property at line 3,

$$\begin{aligned}\mathbb{P}\left((M_1, M_2) \text{ discloses data}\right) &= \mathbb{P}\left(M_1 \text{ or } M_2 \text{ disclose data}\right)\\ &= 1 - \mathbb{P}\left(\text{Neither } M_1 \text{ nor } M_2 \text{ disclose data in the open}\right)\\ &= 1 - (1 - \mathbb{P}\left(M_1 \text{ discloses data}\right))(1 - \mathbb{P}\left(M_2 \text{ discloses data}\right))\\ &\leq 1 - (1-\delta)(1-\delta').\end{aligned}$$

□

One may also prove the previous result by using equation (13) of [10]. Putting everything together,

$$\mathcal{R}\left((M_a, M_b)\right) \overset{?}{\subseteq} \mathcal{R}\left(f_{\varepsilon_1,\delta_1} \otimes f_{\varepsilon_1,\delta_1}\right) \cap \mathcal{R}\left(\varepsilon_1, 1-(1-\delta_1)(1-\delta_2)\right) \cap \mathcal{R}\left(0, 1-(1-\delta_2)^2\right).$$

We compare the hypothesised region to the true one computed from theorem 29 in section 2.5.1 in figure 4.1. Unfortunately, as made obvious by the plot, the computed region is larger than the true one. It seems

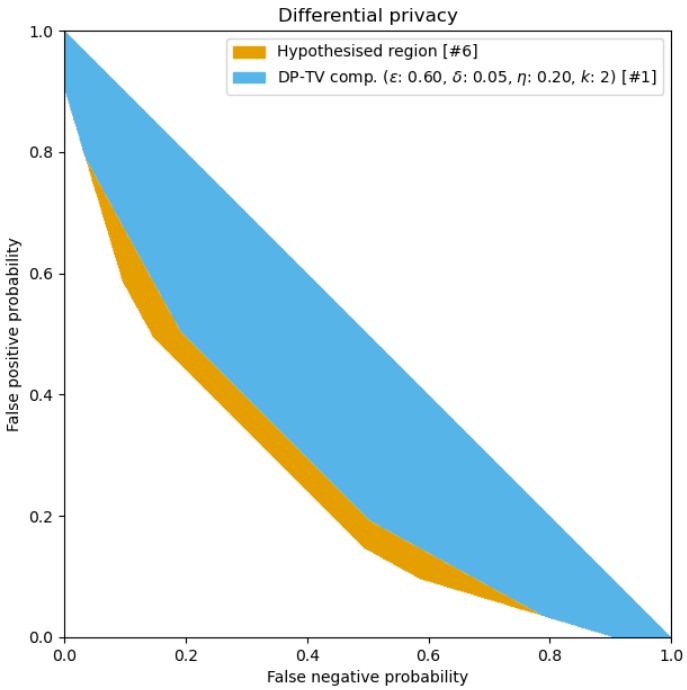

FIGURE 4.1
Comparing the hypothesised region to the true one, with $\varepsilon_1 = 0.6$, $\delta_1 = 0.05$ and $\delta_2 = 0.2$

that the slopes of the intersected DP regions are correct but two of the δ parameters are off. Indeed, when looking at the intersection more in detail, one may realise that the constraint $\mathcal{R}\left(0, 1-(1-\delta_2)^2\right)$ is usually not active. Thus this line of research is still in progress.

4.2 SPLITTING THE INTERSECTION OF MULTIPLE DP CONSTRAINTS INTO A COMPOSITION OF SINGLE ONES

Working on the previous problem, we thought of the following slight generalisation. The authors of [10] show that

$$f_{\varepsilon,\delta} = f_{\varepsilon,0} \otimes f_{0,\delta},$$

decomposing an $(\varepsilon, \delta)-$DP guarantee into two simpler $(\varepsilon, 0)-$DP and $(0, \delta)-$DP constraints. This inspired us to do the same for multiple DP constraints, i.e given $(\varepsilon_1, \delta_1), \dots, (\varepsilon_k, \delta_k) \in \mathbb{R}_+ \times [0, 1]$, finding $m \in \mathbb{N}^*$ and $(\tilde{\varepsilon}_1, \tilde{\delta}_1), \dots, (\tilde{\varepsilon}_m, \tilde{\delta}_m) \in \mathbb{R}_+ \times [0, 1]$ such that

$$f_{(\varepsilon_1,\delta_1),\dots,(\varepsilon_k,\delta_k)} = \bigotimes_{i=1}^{m} f_{\tilde{\varepsilon}_i,\tilde{\delta}_i},$$

where $f_{(\varepsilon_1,\delta_1),\dots,(\varepsilon_k,\delta_k)}$ denotes the ROC curve of the privacy region $\bigcap_{i=1}^{k} \mathcal{R}(\varepsilon_i, \delta_i)$. Then, the right-hand side's composition privacy region can be approximated using theorem 12 in section 2.3.2 or corollary 42 in 3.4.4.

Observe that solving the first problem in this chapter essentially reduces to the second one. Indeed, if $f_{(\varepsilon_1,\delta_1),(\varepsilon_2,\delta_2)} = \bigotimes_{i=1}^{m} f_{\tilde{\varepsilon}_i,\tilde{\delta}_i}$, then

$$f_{(\varepsilon_1,\delta_1),(\varepsilon_2,\delta_2)} \otimes f_{(\varepsilon_1,\delta_1),(\varepsilon_2,\delta_2)} = \left(\bigotimes_{i=1}^{m} f_{\tilde{\varepsilon}_i,\tilde{\delta}_i}\right) \otimes \left(\bigotimes_{i=1}^{m} f_{\tilde{\varepsilon}_i,\tilde{\delta}_i}\right) = \bigotimes_{i=1}^{m} \left(f_{\tilde{\varepsilon}_i,\tilde{\delta}_i}\right)^{\otimes 2},$$

again reducing the problem to the composition of single DP constrained mechanisms. Beyond finding the privacy guarantees for the composition of multi-DP constrained mechanisms, such a decomposition may prove useful in the study of DP theory as it yields an alternative, possibly more suitable way for some contexts, to express the privacy guarantee of a mechanism as a composition. This is especially true when one considers the algebraic framework developed for $f-$DP in [10] as well as the primal-dual perspective on DP, from the same paper, which essentially states that any $f-$DP guarantee can be seen as a (potentially countably infinite) collection of $(\varepsilon_i, \delta_i)-$DP constraints. More specifically:

PROPOSITION 56: PRIMAL-DUAL PERSPECTIVE
For a symmetric trade-off function f, a mechanism is $f-$DP if and only if it is $(\varepsilon, \delta_\varepsilon)-$DP for all $\varepsilon \geq 0$ with

$$\delta_\varepsilon = 1 - \inf_{\alpha \in [0,1]} \alpha e^{\varepsilon} + f(\alpha).$$

Finding where to start in this problem is rather difficult, as determining the smallest m alone is non-trivial. Indeed, reading theorem 11 from 2.3.2 backwards and using the same notation, we find that

$$f_{(2\varepsilon,1-(1-\delta)^2(1-\delta_0)),(0,1-(1-\delta)^2(1-\delta_1))} = f_{\varepsilon,\delta} \otimes f_{\varepsilon,\delta},$$

and

$$f_{(3\varepsilon,1-(1-\delta)^2(1-\delta_0)),(\varepsilon,1-(1-\delta)^2(1-\delta_1))} = f_{\varepsilon,\delta} \otimes f_{\varepsilon,\delta} \otimes f_{\varepsilon,\delta}.$$

Thus the smallest m is a function of not only k, i.e the number of intersected DP constraints, but also ε, δ.

CHAPTER 5

CONCLUSION

SUMMARY

In this thesis, we have created a visualisation tool for differential privacy. It can display and compare parametrised privacy regions, using the binary hypothesis testing interpretation of DP, extending to newer variants of DP such as DP-TV and $f-$DP. Importantly, the tool also shows the privacy regions resulting from DP composition theorems which can be hard to parse otherwise. On top of that, it features the exact privacy regions of multiple common mechanisms along with insightful utilities, as combining the two helps the user understand the privacy-utility trade-off in real applications. In the making of this tool, we have investigated new research directions regarding the composition of DP mechanisms. Specifically, we have looked at the composition of mechanisms for which multiple DP constraints hold, initiating ideas as to how to tackle this problem.

NEXT STEPS

There are multiple paths we can now follow after this project. On the software side, ideas for new features may come up in meetings with scientists or engineers handling sensitive data. Consequently, collaborating with the Center for Digital Trust (C4DT) is to be considered. On the DP side, we can naturally build on the initial ideas presented in this report in chapter 4. Last but not least, on the broader research side, an important measure we wanted to tackle at the beginning of this semester is maximal leakage [11] (ML), its variants (e.g [12], [13], [14]) and its applications (e.g [15], [16], [17]). Throughout multiple meetings, we have developed multiple research ideas involving ML, albeit too far from the core topics of this thesis to make it into this text. Expanding upon these ideas is surely on the agenda.

ACKNOWLEDGEMENTS

As a closing note, I want to sincerely thank everyone in the Mathematics of Information Laboratory for this second semester in the lab. I have learnt so much about privacy, information theory, but also research in general through my semester project and now my Master's thesis, in such a pleasant and open environment.

Of course, special thanks must also go to Yana and Cemre for all of our weekly discussions this semester, which have not only shaped the project bit by bit, but have also given me much inspiration for my upcoming doctoral studies.

APPENDIX A

APPENDIX - CONVENTIONS AND NOTATION

A.1 ACRONYMS

General acronyms:

- LHS = left hand side.
- RHS = right hand side.
- s.t = such that.
- w.l.o.g = without loss of generality.
- w.r.t = with respect to.

Mathematical acronyms:

- cdf = cumulative distribution function.
- conv = convex hull.
- DP = differential privacy.
- i.i.d = independent and identically distributed.
- pmf = probability mass function.
- pdf = probability density function.
- w.p = with probability.

A.2 SETS AND INDICES

- The complement of a set S is $S^C = \Omega \setminus S$ when it is clear from context what the set Ω is.
- The cardinality of a set S is denoted by $\text{card}\,(S) = |S| = \#S$.
- $x_a^b = (x_a, x_{a+1}, \ldots, x_{b-1}, x_b)$.
- $\mathbb{N} = \{0, 1, 2, \ldots\}$.

- $A \subset B \iff (A \subseteq B \wedge A \neq B)$.

A.3 PROBABILITY NOTATION

- $A \sim B$ signifies that A and B follow the same probability distribution.
- $A \perp\!\!\!\perp B$ indicates that A and B are independent.
- $A_1 - A_2 - \ldots - A_n$ indicates that $\{A_i\}_{i=1}^n$ form a Markov chain.
- $\text{Unif}(\mathcal{X})$ denotes the uniform probability distribution on $\mathcal{X}$.
- F_X is the cdf of X.
- P_X is the pmf or pdf of X.

A.4 MISCELLANEOUS

- $\text{sign}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$

APPENDIX B

APPENDIX - ADDITIONAL PROOFS

B.1 THEORETICAL BACKGROUND

B.1.1 INTRODUCING DIFFERENTIAL PRIVACY

PROPOSITION 57: CONSERVATION UNDER POST-PROCESSING
Let $M : \mathcal{X}^n \to \mathcal{Y}$ be $(\varepsilon, \delta)-$DP, and let $F : \mathcal{Y} \to \mathcal{Z}$ be a possibly randomized mapping.
Then $F \circ M$ is $(\varepsilon, \delta)-$DP.

PROOF: Let $X, X' \in \mathcal{X}^n$ and $T \subseteq Y$. F can be considered as a distribution over deterministic functions f. Thus

$$\begin{aligned}
\mathbb{P}\left(F(M(X)) \in T\right) &= \mathbb{E}_{f\sim F}\left(\mathbb{P}\left(f(M(X)) \in T\right)\right) \\
&= \mathbb{E}_{f\sim F}\left(\mathbb{P}\left(M(X) \in f^{-1}(T)\right)\right) \\
&\leq \mathbb{E}_{f\sim F}\left(e^{\varepsilon}\mathbb{P}\left(M(X') \in f^{-1}(T)\right) + \delta\right) \\
&= e^{\varepsilon}\mathbb{P}\left(F(M(X')) \in T\right) + \delta.
\end{aligned}$$

□

COROLLARY 58: GROUP PRIVACY
Let $M : \mathcal{X}^n \to \mathcal{Y}$ be an $(\varepsilon, \delta)-$DP mechanism, and $X, X' \in \mathcal{X}^n$ differing in exactly k positions.
Then for all $T \subseteq \mathcal{Y}$,

$$\mathbb{P}\left(M(X) \in T\right) \leq e^{k\varepsilon}\mathbb{P}\left(M(X') \in T\right) + ke^{k-1}\delta.$$

PROOF: Let $\left\{X^{(i)}\right\}_{i=0}^{k}$ be a sequence of neighboring datasets such that $X^{(0)} = X$ and $X^{(k)} = X'$. Then, using the (ε, δ)-DP property of M,

$$\begin{aligned}
\mathbb{P}\left(M(X^{(0)}) \in T\right) &\leq e^{\varepsilon}\mathbb{P}\left(M(X^{(1)}) \in T\right) + \delta \\
&\leq e^{\varepsilon}(e^{\varepsilon}\mathbb{P}\left(M(X^{(2)}) \in T\right) + \delta) + \delta
\end{aligned}$$

$$
\begin{aligned}
&= e^{2\varepsilon}\mathbb{P}\left(M(X^{(2)}) \in T\right) + \delta(1+e^{\varepsilon}) \\
&\leq e^{2\varepsilon}(e^{\varepsilon}\mathbb{P}\left(M(X^{(3)}) \in T\right) + \delta) + \delta(1+e^{\varepsilon}) \\
&= e^{3\varepsilon}\mathbb{P}\left(M(X^{(3)}) \in T\right) + \delta(1+e^{\varepsilon}+e^{2\varepsilon}) \\
&\vdots \\
&\leq e^{k\varepsilon}\mathbb{P}\left(M(X^{k}) \in T\right) + \delta\sum_{i=0}^{k-1} e^{i\varepsilon}.
\end{aligned}
$$

What is left to observe is that $e^{i\varepsilon} \leq e^{(k-1)\varepsilon}$ for all $i \in [\![0, k-1]\!]$ since $\varepsilon \geq 0$, hence

$$
\sum_{i=0}^{k-1} e^{i\varepsilon} \leq \text{card}\left([\![0, k-1]\!]\right) e^{(k-1)\delta} = ke^{(k-1)\delta}
$$

which concludes the proof. □

B.1.2 COMMON MECHANISMS FOR DIFFERENTIAL PRIVACY

PROPOSITION 59: PRIVACY OF EXPONENTIAL MECHANISM
The exponential mechanism is $(\varepsilon, 0)-$DP.

PROOF: Let X_0, X_1 be two neighbouring datasets.

$$
\begin{aligned}
\frac{\mathbb{P}\left(M(X_0)=c\right)}{\mathbb{P}\left(M(X_1)=c\right)} &= \frac{\exp\left(\frac{\varepsilon s(X_0,c)}{2\Delta}\right)}{\exp\left(\frac{\varepsilon s(X_1,c)}{2\Delta}\right)} \frac{\sum_{c'\in\mathcal{C}}\exp\left(\frac{\varepsilon s(X_1,c')}{2\Delta}\right)}{\sum_{c'\in\mathcal{C}}\exp\left(\frac{\varepsilon s(X_0,c')}{2\Delta}\right)} \\
&= \exp\left(\frac{\varepsilon\left(s(X_0,c)-s(X_1,c)\right)}{2\Delta}\right) \frac{\sum_{c'\in\mathcal{C}}\exp\left(\frac{\varepsilon s(X_1,c')}{2\Delta}\right)}{\sum_{c'\in\mathcal{C}}\exp\left(\frac{\varepsilon s(X_0,c')}{2\Delta}\right)} \\
&\leq \exp\left(\frac{\varepsilon\Delta}{2\Delta}\right)\exp\left(\frac{\varepsilon}{2}\right) \frac{\sum_{c'\in\mathcal{C}}\exp\left(\frac{\varepsilon s(X_0,c')}{2\Delta}\right)}{\sum_{c'\in\mathcal{C}}\exp\left(\frac{\varepsilon s(X_0,c')}{2\Delta}\right)} \\
&\leq \exp\varepsilon.
\end{aligned}
$$

The second to last inequality uses the definition of ℓ_1-sensitivity twice. The case where $\mathcal{C}$ is uncountable is analogous. □

B.2 VISUALISING PRIVACY

B.2.1 COMPUTATION OF SPECIFIC MECHANISMS' PRIVACY REGIONS

PROPOSITION 60: GAUSSIAN MECHANISM TRADE-OFF FUNCTION
Let $f : \mathcal{X}^n \to \mathbb{R}^k$ be a function with finite ℓ_2-sensitivity, $\varepsilon > 0$ and $\delta \in \ ]0, 1]$. Then the Gaussian mechanism as described in 2.4.3 is μ-GDP with $\mu = \frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}$, i.e it is $G_{\frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}}$-DP with

$$G_{\frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}}(\alpha) = \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}\right).$$

PROOF: The proof resembles the Laplace case since both mechanisms have the similar structure $M(X) = f(X) + N$. Thus, let $\Delta = \Delta_{f,2}$ be the ℓ_2-sensitivity of f. Then, letting L be the mechanism's trade-off function,

$$\begin{aligned}
L(\alpha) &= \inf_{\substack{X_0, X_1 \in \mathcal{X}^n \\ X_0, X_1 \text{ neighbouring}}} T\left(\prod_{i=1}^k \mathcal{N}\left(f(X_0)_i, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right), \prod_{i=1}^k \mathcal{N}\left(f(X_1)_i, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right)\right)(\alpha) \\
&= \inf_{\substack{X_0, X_1 \in \mathcal{X}^n \\ X_0, X_1 \text{ neighbouring}}} T\left(\prod_{i=1}^k \mathcal{N}\left(f(X_0)_i - f(X_1)_i, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right), \prod_{i=1}^k \mathcal{N}\left(0, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right)\right)(\alpha) \\
&= T\left(\mathcal{N}\left(\Delta, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right), \mathcal{N}\left(0, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right)\right)(\alpha) \\
&= T\left(\mathcal{N}\left(\Delta, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right), \mathcal{N}\left(0, 2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}\right)\right)(\alpha) \\
&= T\left(\sqrt{2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}}\mathcal{N}\left(\frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}, 1\right), \sqrt{2\log\left(\frac{5}{4\delta}\right)\frac{\Delta^2}{\varepsilon^2}}\mathcal{N}(0,1)\right)(\alpha) \\
&= T\left(\mathcal{N}\left(\frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}, 1\right), \mathcal{N}(0,1)\right)(\alpha) \\
&= G_{\frac{\varepsilon}{\sqrt{2\log\left(\frac{5}{4\delta}\right)}}}(\alpha).
\end{aligned}$$

□

LEMMA 61: OPTIMAL TEST REGION FOR THE EXPONENTIAL MECHANISM
Let M be the exponential mechanism on $\mathcal{X}^n$ with score function $s : \mathcal{X}^n \times \mathcal{C} \to \mathbb{R}$ and DP parameter $\varepsilon > 0$. Let X_0, X_1 be neighbouring datasets in $\mathcal{X}^n$. Fix a maximum false positive error rate $\alpha \in]0, 1[$, i.e, for a test region $T \subseteq \mathcal{C}$, require that

$$\text{FP}\left(X_0, X_1, M, T\right) = \mathbb{P}\left(M(X_0) \in T\right) \leq \alpha.$$

Then, there exists $t_\alpha(X_0, X_1) \in \mathbb{R}$ such that the optimal T, minimizing the false-negative probability $\text{FN}\left(X_0, X_1, M, T\right) = \mathbb{P}\left(M(X_1) \in T^C\right)$, is

$$T = \left\{c \in \mathcal{C} \mid s(X_1, c) - s(X_0, c) \geq t_\alpha(X_0, X_1)\right\}.$$

PROOF: Fixing some notation first: we write the pmf or pdf of the exponential mechanism's output as follows,

$$P_{M(X)}(c) = \frac{1}{\mu_X} \exp\left(\frac{\varepsilon}{2\Delta} s(X, c)\right)$$

where Δ is the used upper-bound to the ℓ_1-sensitivities of $s(\cdot, c)$.

By the Neyman-Pearson Lemma, there exists $\eta_\alpha > 0$ such that

$$T = \left\{c \in \mathcal{C} \mid \frac{P_{M(X_1)}(c)}{P_{M(X_0)}(c)} \geq \eta_\alpha\right\}.$$

It thus suffices to simplify the likelihood ratio.

$$\begin{aligned}
\frac{P_{M(X_1)}(c)}{P_{M(X_0)}(c)} \geq \eta_\alpha &\iff \frac{\exp\left(\frac{\varepsilon}{2\Delta} s(X_1, c)\right)}{\exp\left(\frac{\varepsilon}{2\Delta} s(X_0, c)\right)} \geq \eta_\alpha \frac{\mu_{X_1}}{\mu_{X_0}} \\
&\iff \exp\left(\frac{\varepsilon}{2\Delta}\left(s(X_1, c) - s(X_0, c)\right)\right) \geq \eta_\alpha \frac{\mu_{X_1}}{\mu_{X_0}} \\
&\iff s(X_1, c) - s(X_0, c) \geq \frac{2\Delta}{\varepsilon} \log\left(\eta_\alpha \frac{\mu_{X_1}}{\mu_{X_0}}\right).
\end{aligned}$$

Setting $t_\alpha(X_0, X_1) = \frac{2\Delta}{\varepsilon} \log\left(\eta_\alpha \frac{\mu_{X_1}}{\mu_{X_0}}\right)$ concludes the proof. □

B.2.2 VISUALISING PRIVACY-UTILITY TRADE-OFFS

PROPOSITION 62:
In the same setup as the previous definition,

$$\text{MSE}(\text{Mean}, \text{DPMean}_{\varepsilon,\delta}) = 2k \log\left(\frac{5}{4\delta}\right) \frac{d^2}{n^2\varepsilon^2}.$$

PROOF:

$$\text{MSE}(\text{Mean}, \text{DPMean}_{\varepsilon,\delta}) = \mathbb{E}\left(||\text{Mean} - \text{Mean} - (Y_i)_{i=1}^k||_2^2\right)$$

$$
\begin{aligned}
&= \mathbb{E}\left(||\left(Y_i\right)_{i=1}^{k}||_2^2\right) \\
&= k\text{Var}\left(\mathcal{N}\left(0, 2\log\left(\frac{5}{4\delta}\right)\frac{d^2}{n^2\varepsilon^2}\right)\right) \\
&= 2k\log\left(\frac{5}{4\delta}\right)\frac{d^2}{n^2\varepsilon^2}.
\end{aligned}
$$

□

LEMMA 63:
The median's exponential mechanism score function q has sensitivity $\Delta \leq 2$.

PROOF: Let $X_0, X_1 \in \mathcal{X}^n$ be neighbouring datasets. Assume w.l.o.g that $X_{0,1} < X_{1,1}$, and $X_{0,j} = X_{1,j}$ for $j \in [\![2, n]\!]$. Let $x \in \mathcal{X}$. If $x < X_{0,1}$ or $x > X_{1,1}$, then $\text{sign}\left(x - X_{0,1}\right) = \text{sign}\left(x - X_{1,1}\right)$, hence $q(X_0, x) = q(X_1, x)$ - as the other terms in the summation defining q do not change. If $x = X_{0,1}$ then $\text{sign}\left(x - X_{0,1}\right) = 0$ and $\text{sign}\left(x - X_{1,1}\right) = -1$, thus $|q(X_0, x) - q(X_1, x)| = 1$. The case $x = X_{1,1}$ is similar. Lastly, if $x \in [\![X_{0,1}, X_{1,1}]\!]$, then $\text{sign}\left(x - X_{0,1}\right) = -1$ and $\text{sign}\left(x - X_{1,1}\right) = 1$, thus $|q(X_0, x) - q(X_1, x)| = 2$.
Note that the bound 2 can only be attained for $n, m \geq 3$, for some $x \in [\![X_{0,1}, X_{1,1}]\!]$ to possibly exist. □

BIBLIOGRAPHY

[1] † LiudasPanavas et al. 'A Visualization Tool to Help Technical Practitioners of Differential Privacy *'. In: URL: https://api.semanticscholar.org/CorpusID:272200073.

[2] Gautham Kamath. *Algorithms for Private Data Analysis*. 2020. URL: http://www.gautamkamath.com/CS860-fa2020.html.

[3] Cynthia Dwork et al. 'Calibrating Noise to Sensitivity in Private Data Analysis'. In: *J. Priv. Confidentiality* 7 (2006), pp. 17–51. URL: https://api.semanticscholar.org/CorpusID:2468323.

[4] Larry Wasserman and Shuheng Zhou. *A statistical framework for differential privacy*. 2009. arXiv: 0811.2501 [math.ST]. URL: https://arxiv.org/abs/0811.2501.

[5] Peter Kairouz, Sewoong Oh and Pramod Viswanath. *The Composition Theorem for Differential Privacy*. 2015. arXiv: 1311.0776 [cs.DS]. URL: https://arxiv.org/abs/1311.0776.

[6] Cynthia Dwork and Aaron Roth. 2014. DOI: 10.1561/0400000042.

[7] Ilya Mironov. 'Rényi Differential Privacy'. In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. 2017, pp. 263–275. DOI: 10.1109/CSF.2017.11.

[8] Mengmeng Yang et al. 'Local differential privacy and its applications: A comprehensive survey'. In: *Computer Standards & Interfaces* 89 (2024), p. 103827. ISSN: 0920-5489. DOI: https://doi.org/10.1016/j.csi.2023.103827. URL: https://www.sciencedirect.com/science/article/pii/S0920548923001083.

[9] Elena Ghazi and Ibrahim Issa. *Total Variation Meets Differential Privacy*. 2024. arXiv: 2311.01553 [cs.IT]. URL: https://arxiv.org/abs/2311.01553.

[10] Jinshuo Dong, Aaron Roth and Weijie J. Su. *Gaussian Differential Privacy*. 2019. arXiv: 1905.02383 [cs.LG]. URL: https://arxiv.org/abs/1905.02383.

[11] Ibrahim Issa, Aaron B. Wagner and Sudeep Kamath. 'An Operational Approach to Information Leakage'. In: *IEEE Transactions on Information Theory* 66.3 (2020), pp. 1625–1657. DOI: 10.1109/TIT.2019.2962804.

[12] Robinson D. H. Cung, Yanina Y. Shkel and Ibrahim Issa. 'Binary Maximal Leakage'. In: *2024 IEEE International Symposium on Information Theory (ISIT)*. 2024, pp. 2748–2753. DOI: 10.1109/ISIT57864.2024.10619387.

[13] Sara Saeidian et al. 'Pointwise Maximal Leakage'. In: *IEEE Transactions on Information Theory* 69.12 (2023), pp. 8054–8080. DOI: 10.1109/TIT.2023.3304378.

[14] Shuaiqi Wang, Zinan Lin and Giulia Fanti. 'Statistic Maximal Leakage'. In: *2024 IEEE International Symposium on Information Theory (ISIT)*. 2024, pp. 2742–2747. DOI: 10.1109/ISIT57864.2024.10619258.

[15] Ibrahim Issa, Amedeo Roberto Esposito and Michael Gastpar. *Generalization Error Bounds for Noisy, Iterative Algorithms via Maximal Leakage*. 2023. arXiv: 2302.14518 [cs.LG]. URL: https://arxiv.org/abs/2302.14518.

[16] Benjamin Wu, Aaron B. Wagner and G. Edward Suh. *A Case for Maximal Leakage as a Side Channel Leakage Metric*. 2020. arXiv: 2004.08035 [cs.IT]. URL: https://arxiv.org/abs/2004.08035.

[17] Hatef Otroshi Shahreza, Yanina Y. Shkel and Sébastien Marcel. 'Measuring Linkability of Protected Biometric Templates Using Maximal Leakage'. In: *IEEE Transactions on Information Forensics and Security* 18 (2023), pp. 2262–2275. DOI: 10.1109/TIFS.2023.3266170.