Extending the Foundations of Differential Privacy: Robustness and Flexibility

Anonymous Author(s)
Affiliation
Address
email

Abstract

Differential Privacy (DP) is an area that has recently seen many direct and indirect applications to machine learning. In this work, we make foundational contributions to the area of DP.

Our first contribution is definitional. We define two complementary concepts that greatly enhance the applicability of DP, namely, robust privacy and flexible accuracy. Robust privacy requires that a mechanism provides the "best possible privacy" without further degrading accuracy guarantees, even if such privacy is not a priori anticipated based on input neighborhoods alone. Flexible accuracy allows small distortions in the input (e.g., dropping outliers) before measuring accuracy of the output. Along the way, we also extend the notion of DP to sampling (i.e. computation of randomized functions).

Our second contribution is in establishing versatile composition theorems that relate these notions.

Our third contribution is constructive: We present mechanisms that can help in achieving these notions, where previously no meaningful differentially private mechanisms were possible. In particular, we illustrate an application to differentially private histograms, which in turn yields mechanisms for revealing the support of a dataset or the extremal values in the data.

1 Introduction

Blurb to connect DP and ML.

Below, we identify and address two limitations of the DP framework that seem to have evaded attention. At a high-level, these limitations follow from a seemingly natural choice: Accuracy guarantees of a mechanism are in terms of distances in the output space, and privacy demands are in terms of distances in the input space (neighboring inputs). Somewhat surprisingly, these choices turn out to be not always adequate. Our extensions can be seen as adding accuracy guarantees in terms of distances (or rather, distortions) in the input space, and privacy demands in terms of distances in the output space. Along the way, we extend the notion of DP to randomized functions over a metric space, for which distances are measured using a (generalization of) Wasserstein distance. These extensions greatly expand the scope of DP. Apart from the direct implications to privacy of training data, we anticipate that the implications of DP to generalization guarantees (as shown recenty in cite) will also be strengthened by these extensions.

We start by discussing the limitations of DP.

Lack of Flexibility. Consider a simplistic learning task which tries to learn an upper bound on integer valued observations – say, ages of patients who recovered from a certain disease – presented to

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

it. For the sake of privacy, one may wish to apply a DP mechanism, rather than output the maximum in the sample itself. Two possible datasets which differ in only one patient are considered neighbors and a DP mechanism needs to make the outputs on these two samples indistinguishable from each other. However, the function in question is *highly sensitive* – two neighboring datasets can have their maxima differ by as much as the entire range of possible ages – and the standard DP mechanisms in the literature will add so much noise that no useful information can be retained.[1]

As we shall see, the above limitation can be attributed to a rigidly defined notion of accuracy. This same rigidity leads to another surprising limitation too. Consider the problem of reporting a *histogram* (again, say, of patients' ages). Here a standard DP mechansim, of adding a zero-mean Laplace noise to each bar of the histogram is indeed reasonable, as the histogram function has low sensitivity in each bar. Now, note that *maximum can be computed as a function of the histogram*. However, even though the histogram mechanism was sufficiently accurate in the standard sense, the maximum computed from its output is no longer accurate! This is because when a non-zero count is added to a large-valued item which originally has a count of 0, the maximum can increase arbitrarily.

In this work we develop a more relaxed notion of accuracy, called *flexible accuracy*, that lets us address both of the above issues. In particular, it not only enables new DP mechanisms for maximum, but also allows one to derive the mechanism from a new DP mechanism for histograms. A composition theorem enables us to transfer the accuracy guarantees on histogram to accuracy guarantees on the maximum function.

Lack of Robustness. Differential Privacy focuses on making outputs from *neighboring* databases indistinguishable, where neighborhood usually refers to databases obtained by adding or deleting a small number of data items (or a single one). However, such a notion of neighborhood of the databases may not capture all pairs of databases that *should be* indistinguishable from each other.

Consider training a machine learning model on either dataset D_1 or dataset D_2, where the two datasets are disjoint. Suppose both the datasets are representative and yield very similar models. In this case, we may reasonably require that querying a model should not reveal whether it was trained on D_1 or D_2. Indeed, since the models are "similar," one may expect them to yield results which are indistinguishable from each other. Unfortunately, this is not generally true: Similarity of outputs is measured in terms of a distance in the output space (or rather, the Wasserstein distance over that space, since the output is probabilistic); but the extent of their indistinguishability is measured in terms of total variation distance or the ratio of probabilities (as in DP), which are not influenced by the metric space associated with the outputs. For instance, if the output from the model trained in D_1 has an even value for the least significant digit, and the other has an odd value, the total variation distance between the two output distributions is maximum, while the Wasserstein distance can be very small.

In short, DP only guarantees indistinguishability between datasets which are close to each other in the input space, whereas one may demand – without necessarily compromising on accuracy – indistinguishability between datasets which result in outputs that are close to each other. Robustness is a complementary notion defined for a mechanism that addresses this.

1.1 Our Contributions

blurb

Felxibility. Flexible accuracy is a notion that is designed to salvage the situation for functions like maximum. The high-level idea is to allow for some *distortion of the input* when measuring accuracy. We shall require distortion to be defined using a *quasi-metric* over the input sapce (a quasi-metric is akin to a metric, but is not required to be symmetric). A typical form of distortion is to *drop a few items* from the dataset; in this case, adding a data item is not considered low distortion. Referring back to the example of reporting maximum, given a dataset with a single elderly patient and many young patients, flexible accuracy with respect to this distortion allows a mechanism for maximum to report the maximum age of the younger group.

[1] Indeed, *all datasets* with low maximum values have high sensitivity *locally*, by considering a neighboring dataset with a single additional data item with a large value. As such, mechanisms which add noise based on the local sensitivity rather than global sensitivity cite also do not fare any better.

Flexible accuracy also provides us with a means for *transferring accuracy guarantees* when composed with other functions or mechanisms. Consider again the example of the histogram and maximum functions from above. Recall that even a high (but less than perfect) accuracy of histograms under a metric in the output space can result in maximum computed from the histogram to be wildly inaccurate. But if the inaccuracy in the histogram can be entirely attributed to a distortion in the input, computing maximum on this histogram does not amplify the inaccuracy at all.

Robustness. We define a mechanism whose output is in a metric space to be robust if, roughly, it holds that whenever two input distributions result in output distributions that are close in Wasserstein distance, then the output distributions are also indistinguishable in the sense of differential privacy. Unlike in the definition of differential privacy, where an input neighborhood is specified, here the neighborhood is implicitly defined by the mechanism itself.

Composition Theorems.

Constructions.

Related Work. Book [1]

2 Preliminaries

Notations. We denote by $\mathbb{N}$ all the non-negative integers (including zero). For $i, j \in \mathbb{N}$, such that $i \leq j$, we write $[i : j]$ to denote the set $\{i, i+1, \ldots, j\}$.

Definition 1 (Total Variation Distance)**.** *Let p and q be two discrete probability distributions on a sample space Ω. The total variation distance between p and q, denoted by $\Delta(p,q)$, is defined as follows:*

$$\Delta(p,q) = \frac{1}{2}\sum_{\omega\in\Omega}|p(w) - q(w)|.$$

Let $\theta \in [0,1]$ be a constant, and let $\Phi^\theta(p,q)$ denote the set of all joint distributions ϕ^θ with marginals ϕ_1^θ and ϕ_2^θ such that $\Delta(\phi_1^\theta, p) + \Delta(\phi_2^\theta, q) \leq \theta$ hold. Note that at $\theta = 0$, all the joint distributions in $\Phi^0(p,q)$ have marginals exactly equal to p and q. In this case we will write $\Phi(p,q)$ to denote $\Phi^0(p,q)$.

Definition 2 (Wasserstein Distance)**.** *Let p and q be two discrete probability distributions on $\mathbb{R}$. The Wasserstein distance between p and q is defined as:*

$$W(p,q) = \inf_{\phi\in\Phi(p,q)} \mathbb{E}_{(x,y)\leftarrow\phi}[|x-y|]. \tag{1}$$

Definition 3 (θ-Wasserstein Distance)**.** *Let p and q be two discrete probability distributions on $\mathbb{R}$. Let $\theta \in [0,1]$. The θ-Wasserstein distance between p and q is defined as:*

$$W^\theta(p,q) = \inf_{\phi\in\Phi^\theta(p,q)} \mathbb{E}_{(x,y)\leftarrow\phi}[|x-y|]. \tag{2}$$

Definition 4 (∞-Wasserstein Distance)**.** *Let p and q be two discrete probability distributions on $\mathbb{R}$. The ∞-Wasserstein distance between p and q is defined as:*

$$W_\infty(p,q) = \inf_{\phi\in\Phi(p,q)} \max_{(x,y)\leftarrow\phi} |x-y|. \tag{3}$$

Definition 5 ((∞,θ)-Wasserstein Distance)**.** *Let p and q be two discrete probability distributions on $\mathbb{R}$. Let $\theta \in [0,1]$. The (∞,θ)-Wasserstein distance between p and q is defined as:*

$$W_\infty^\theta(p,q) = \inf_{\phi\in\Phi^\theta(p,q)} \max_{(x,y)\leftarrow\phi} |x-y|. \tag{4}$$

Claim 1. *Let X and Y be a random variable with distributions μ_1, μ_2, respectively. Let Z be a an independent noise random variable with a distribution μ_3. Then we have,*

$$\begin{aligned} W^\gamma(X,Y) &= W^\gamma(X+Z, Y+Z), \\ W_\infty^\gamma(X,Y) &= W_\infty^\gamma(X+Z, Y+Z), \end{aligned}$$

i.e, convolution of distributions with the same independent noise does not change Wasserstein distance.

Proof. First, we show that applying convolution of distributions with another does not increase the Wasserstein distance between them.

Let p, q, r be three distributions, with wasserstein distance between p,q as β and π be the optimal transfer which achieves this . Let I_1, I_2 be two distributions obtained by convolving p, q with r respectively. i.e,

$$Pr(I_1 = x) = \int_Z p(x-z)r(z)$$
$$Pr(B + C = y) = \int_Z q(y-z)r(z)$$

From the above equations, we can construct a transfer from I_1, I_2 as follows. Let z be drawn from r with a probability $p(z)$. In this event, we transfer I_1 to I_2 using the policy, π + z. This policy also transfers wasserstein distance of β for all z. Hence, it's expectation over z is also β. Since, wasserstein distance is an infimum of all transfers, we have, the wasserstein distance between I_1, I_2 to be at most β. Hence, proved.

Now, for the random variables X, Y, Z. The distributions of X + Z, Y + Z are given by the following convolutions.

$$Pr(A + C = x) = \int_Z \mu_1(x-z)\mu_3(z)$$
$$Pr(B + C = y) = \int_Z \mu_2(y-z)\mu_3(z)$$

From the above reasoning, we conclude that the wasserstein distance between X + Z, Y + Z is atmost that of X, Y.

Note that the distributions of X and Y can be obtained by inverse convolution of Z with that of X + Z, Y + Z. This is equivalent to convolution with an appropriate distribution Z'. This is true for discrete variables, convolution is matrix multiplication with a kernel, so inverse convolution is just multiplying with inverse of the kernel. How does it go for continuous variables ? . Again, as convolving with a distribution doesn't increase the wasserstein distance, we conclude the wasserstein distance between X + Z, ,Y + Z is atmost that of X, Y.

Combining both these results, we have that, wasserstein distance between X, Y is same as wasserstein distance between X + Z, Y + Z. ☐

Claim 2. *We have the following triangle inequality for the θ-Wasserstein distance, where p, q, r are any three distributions defined over the same support, and $\gamma_1, \gamma_2 \geq 0$.*

$$W_\infty^{\gamma_1+\gamma_2}(p,r) \leq W_\infty^{\gamma_1}(p,q) + W_\infty^{\gamma_2}(q,r).$$
$$W^{\gamma_1+\gamma_2}(p,r) \leq W^{\gamma_1}(p,q) + W^{\gamma_2}(q,r).$$

Proof. We will prove the result for W_∞. The result for W is exactly the same.

We first prove the following result. Let p,q be two given distribution. Let p' be a distribution which is δ statistical distance apart from p. Then there exists another distribution q', which is atmost δ statistical distance apart from q, such that

$$W_\infty(p', q) = W_\infty(p, q')$$

Intuitively, this can be proved as follow. Let ϕ be any joint distribution of p' and q. We will convert this joint distribution into another joint distribution ϕ' whose marginals are p and q' such that $\max_{(x,y)\leftarrow\phi} |x-y| = \max_{(x,y)\leftarrow\phi'} |x-y|$. p' has extra mass at some places and less mass at other places as compared to p. In ϕ, this extra mass (which is equal to δ) would be joined with some δ mass of q. Now we restore p by moving this δ extra part so as to convert p' into p but we also move the corresponding δ mass of q parallelly (keeping the distance same). The resulting joint distribution is the required ϕ'. Its easy to see that the total mass of q moved is δ so the q' we get has atmost δ statistical difference with q. And since we moved the masses of p' and q parallely, no distance changed. This implies $\max_{(x,y)\leftarrow\phi} |x-y| = \max_{(x,y)\leftarrow\phi'} |x-y|$.

This allows us to say that $W_\infty(p', q) \geq W_\infty(p, q')$. We can use argument to prove that $W_\infty(p', q) \leq W_\infty(p, q')$. Hence we proved that $W_\infty(p', q) \geq W_\infty(p, q')$.

Let us use $W_{\infty}^{\gamma_a,\gamma_b}(p,q)$ to represent $\inf_{\substack{p',q': \\ \Delta(p,p')\le\gamma_a, \\ \Delta(q,q')\le\gamma_b}} W_{\infty}(p',q')$. The above result implies that, $W_{\infty}^{\gamma}(p,q)$ is equal to $W_{\infty}^{\gamma_a,\gamma_b}(p,q) \forall \gamma_a, \gamma_b such that \gamma_a + \gamma_b = \gamma$.

Now using our above proof, we can say that

$$\begin{aligned} W_{\infty}^{\gamma_1+\gamma_2}(p,r) &= W_{\infty}^{\gamma_1,\gamma_2}(p,r) \\ &\le W_{\infty}^{\gamma_1,0}(p,q) + W_{\infty}^{0,\gamma_2}(q,r) && \text{(since wasserstein distance is a metric)} \\ &= W_{\infty}^{\gamma_1}(p,q) + W_{\infty}^{\gamma_2}(q,r) \end{aligned}$$

□

Let $\mathcal{X}$ be a finite set, and let $\mathbf{x} := (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ represent a database with n elements, where $x_i \in \mathcal{X}$ is the i-th element. For any two databases $\mathbf{x}, \mathbf{x}' \in \mathcal{X}^n$, we say that $\mathbf{x} \sim \mathbf{x}'$, if $\mathbf{x}'$ is obtained from $\mathbf{x}$ by adding/removing a single element. Let $f : \mathcal{X}^n \to \mathcal{Y}$ be a randomized function, where $\mathcal{Y} = \{y_1, y_2, \ldots, y_k\}$ is a discrete set for some finite $k \in \mathbb{N}$. For every $\mathbf{x} \in \mathcal{X}^n$, we have that $\sum_{i=1}^{k} \Pr[f(\mathbf{x}) = y_i] = 1$. **our results are also for continuous functions.**

Definition 6 (Parameterized Sensitivity of a Randomized Query)**.** *For $\theta \in [0,1]$, we define θ-sensitivity of a randomized query f, denoted by $S^{\theta}(f)$, as follows:*

$$S^{\theta}(f) := \max_{\substack{\mathbf{x},\mathbf{x}' \in \mathcal{X}^n: \\ \mathbf{x}\sim\mathbf{x}'}} W_{\infty}^{\theta}(f(\mathbf{x}), f(\mathbf{x}')). \tag{5}$$

Definition 7 (Laplace Distribution)**.** *Let b be a positive real number. Laplace distribution with respect to b (b is called the scaling parameter) and mean μ, denoted by $Lap(x|\mu, b)$, is defined as:*

$$Lap(x|\mu,b) := \frac{1}{2b} e^{\frac{-|x-\mu|}{b}}, \quad x \in \mathbb{R}.$$

We denote a random variable that is distributed as the Laplace distribution with the scaling parameter b and mean μ by $Lap(b,\mu)$. If mean μ is zero, then we will simply denote it by $Lap(b)$.

Definition 8 ((α,β,γ)-accuracy)**.** *Let $d : \mathcal{X}^n \times \mathcal{X}^n \to [0,\infty)$ be a distortion function, not necessarily a distance metric. A mechanism $\mathcal{M}$ for computing a given randomized function $f : \mathcal{X}^n \to \mathcal{Y}$ is said to be (α,β,γ)-accurate with respect to d, if for every $\mathbf{x} \in \mathcal{X}^n$, there exists a random variable X', which is a distribution on all $\mathbf{x}'$ such that $d(\mathbf{x},\mathbf{x}') \le \alpha$ and that $W^{\gamma}(f(X'), \mathcal{M}(\mathbf{x})) \le \beta$.*

Definition 9 (Robustness)**.** *Let $\rho, \theta, \epsilon, \delta$ be non-negative real numbers with $\delta, \theta \le 1$. Let $\mathcal{RM} : A \to B$ be a randomized mechanism and p_x denote the output distribution of $\mathcal{RM}$ when the input is taken from a distribution x over A. The mechanism $\mathcal{RM}$ is called $(\theta,\rho,\epsilon,\delta)$-robust if $\forall x, x', S$ where $S \subseteq \mathbb{R}$ and x and x' are two distributions over A such that $W_{\infty}^{\theta}(p_x, p_{x'}) \le \rho$,*

$$\Pr_{y \leftarrow p_x}[y \in S] \le e^{\epsilon} \Pr_{y \leftarrow p_x'}[y \in S] + \delta \tag{6}$$

Definition 10 (distortion sensitivity)**.** *Let $M : A \to B$ be a mechanism for computing a function f and d_1, d_2 are distortion functions on A, B, respectively. We define distortion-sensitivity function σ_f for the randomized function f w.r.t. d_1, d_2 as $\sigma_f(\alpha)$ to be the minimum number such that for every $x \in A$ and $y \in B$, s.t $d_2(f(x), y) \le \alpha$, there exists an $x' \in A$ such that $d_1(x, x') \le \sigma_f(\alpha)$ and $f(x') = y$.*

Note that this may not be well defined for every randomized function.

Definition 11 (θ-error sensitivity)**.** *Let $M : A \to B$ be a mechanism for computing a function f and d_1, d_2 are distance functions on A, B, respectively. Additionally, suppose that the input to the mechanism M is drawn from a distribution. We define θ-error-sensitivity function τ for the mechanism M w.r.t d_1, d_2 as*

$$\tau_M^{\theta}(\beta) = \max_{\substack{\mathbf{x}\sim p, \mathbf{x}'\sim q: \\ W^{\theta}(p,q)\le\beta}} .W^{\theta}(M(\mathbf{x}), M(\mathbf{x}')) \tag{7}$$

3 Adding Robustness using Robust Mechanisms for Identity Functions

3.1 Robust Mechanism for Identity Function over $\mathbb{R}$

We will now show a mechanism $\mathcal{M}_{\text{rob}} : \mathbb{R} \to \mathbb{R}$ which is a robust mechanism for identity function, i.e., the output corresponding to any input is the same as input. Now Suppose the input distribution is x. Let ρ be any non-negative real number and ϵ be any positive real number. Consider the following mechanism for the identity function:

Mechanism 1. $\mathcal{M}_{\text{rob}}$*: On input y, sample z according to the probability distributions $Lap(\frac{\rho}{\epsilon})$ and output $y + z$.*

Lemma 1. *For a fixed constant $\theta \in [0, 1]$, $\mathcal{M}_{\text{rob}}$ achieves $(\theta, \rho, \epsilon, \delta)$-robustness where $\delta = 2e^{\epsilon}\theta$.*

Proof. Fix a constant $\theta \geq 0$. Suppose q and q' denote two input distributions, and let p and p' denote the corresponding output probability distributions of $\mathcal{M}_{\text{rob}}$. Suppose $W^{\theta}_{\infty}(p, p') \leq \rho$. Then by Claim 1, $W^{\theta}_{\infty}(q, q') \leq \rho$. Let $b := \frac{\rho}{\epsilon}$ and $Z \sim Lap(b)$. We need to show that $\forall S \subseteq \mathbb{R}$, $\Pr_{y \leftarrow p}[y \in S] \leq e^{\epsilon} \Pr_{y \leftarrow p'}[y \in S] + \delta$.

$$\Pr_{y \leftarrow p}[y \in S] \;=\; \int_{t \in S} p(t) \cdot dt \;=\; \int_{t \in S} \big[\int_{-\infty}^{\infty} q(s) \cdot p_Z(t - s) \cdot ds \big] \cdot dt \tag{8}$$

Let $\phi^{\theta} \in \Phi^{\theta}(q, q')$ be a joint distribution. Let M^{ϕ} denote $\max_{(x,y) \leftarrow \phi^{\theta}} |x - y|$ for the distribution ϕ^{θ}. For $i \in \mathbb{R}$, let $\theta_1(i) := |\phi^{\theta}_1(i) - q(i)|$ and $\theta_2(i) := |\phi^{\theta}_2(i) - q'(i)|$. Note that, by definition of the distribution ϕ^{θ}, we have $\Delta(\phi^{\theta}_1, q) + \Delta(\phi^{\theta}_2, q') \leq \theta$ which implies $\int_{-\infty}^{\infty} \theta_1(i) \cdot di + \int_{-\infty}^{\infty} \theta_2(i) \cdot di \leq \theta$. Using these in (8) we get the following:

$$\begin{aligned}
\Pr_{y \leftarrow p}[y \in S] &\leq \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} (\phi^{\theta}_1(i) + \theta_1(i)) \cdot e^{\frac{-|t-i|}{b}} \cdot di\big] \cdot dt \\
&= \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} \phi^{\theta}_1(i) \cdot e^{\frac{-|t-i|}{b}} \cdot di\big] \cdot dt + \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} \theta_1(i) \cdot e^{\frac{-|t-i|}{b}} \cdot di\big] \cdot dt \\
&= \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} \phi^{\theta}_1(i) \cdot e^{\frac{-|t-i|}{b}} \cdot di\big] \cdot dt + \int_{-\infty}^{\infty} \theta_1(i) \cdot \underbrace{\int_{t \in S} \frac{1}{2b} e^{\frac{-|t-i|}{b}} \cdot dt}_{\leq 1} \cdot di
\end{aligned} \tag{9}$$

By properties of joint distributions, we have $\phi^{\theta}_1(i) = \int_{-\infty}^{\infty} \phi^{\theta}(i, j) \cdot dj$ and $\phi^{\theta}_2(j) = \int_{-\infty}^{\infty} \phi^{\theta}(i, j) \cdot di$; and by triangle inequality we have $|t - i| \geq |t - j| - |i - j|$. Let $\int_{-\infty}^{\infty} \theta_l(i) \cdot di$ be ω_l, $l \in \{1, 2\}$. Substituting all these in (9) we get the following:

$$\Pr_{y \leftarrow p}[y \in S] \leq \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} \big[\int_{-\infty}^{\infty} \phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|+|i-j|}{b}} \cdot dj\big] \cdot di\big] \cdot dt + \omega_1. \tag{10}$$

Observe that $\phi^{\theta}(i, j)$ is non-zero only if $|i - j| \leq M^{\phi}$ (by definition of M^{ϕ}). Using this for every $i, j \in \mathbb{R}$, the integrand in (10) can be upper-bounded as follows:

$$\phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|}{b}} e^{\frac{|i-j|}{b}} \leq \phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|}{b}} e^{\frac{M^{\phi}}{b}} = \phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|}{b}} e^{\frac{M^{\phi}}{b}}.$$

Substituting this in (10) and with some algebraic manipulations done in Appendix B, we get $\Pr_{y \leftarrow p}[y \in S] \leq e^{\epsilon} \Pr_{y \leftarrow p'}[y \in S] + 2\theta e^{\epsilon}$. This completes the proof of Lemma 1. $\square$

Lemma 2. *For every constant $\gamma' \geq 0$, $\mathcal{M}_{\text{rob}}$ is $(0, \beta', \gamma')$-accurate, where $\beta' = \frac{\rho}{\epsilon(1-\gamma')} \Big(1 - \gamma'[1 + \ln(\frac{1}{\gamma'})]\Big)$. Note that if $\gamma' = 0$, $\beta' = \frac{\rho}{\epsilon}$.*

Proof. Fix a constant $\gamma \geq 0$ and any input x $\in \mathbb{R}$. Instead of treating x as a real, for this proof, we will treat x as a point distribution over the input space from which we are sampling the inputs. Clearly, this is equivalent to treating x as a deterministic input. Let q denote the output distribution of

$\mathcal{M}_{\text{rob}}$ when the input is drawn from x. Let b denote $\frac{\rho}{\epsilon}$. We want to show that $W^\gamma(x, q) \le \beta$, for the above-mentioned β. By definition of W^γ from (2) we have $W^\gamma(x, q) = \inf_{\phi \in \Phi^\gamma(x,q)} \mathbb{E}_{(y,t)\leftarrow\phi}[|y - t|]$.

Consider the following ϕ^*:

$$\phi^*(i, t) = \begin{cases} 0 & \text{if } t < -b\ln(\frac{1}{\gamma}) + i \text{ or } t > b\ln(\frac{1}{\gamma}) + i; \\ \frac{1}{1-\gamma} Lap(t|b, i)x(i) & \text{if } t \in [-b\ln(\frac{1}{\gamma}) + i, b\ln(\frac{1}{\gamma}) + i]. \end{cases} \tag{11}$$

It can be verified that $\Delta(\phi_1^*, x) = 0$ and $\Delta(\phi_2^*, q) \le \gamma$, which implies that $\phi^* \in \Phi^\gamma(x, q)$. This in turn implies that $W^\gamma(x, Q) \le \mathbb{E}_{(y,t)\leftarrow\phi^*}[|y - t|]$. We show in Appendix B that $\mathbb{E}_{(y,t)\leftarrow\phi^*}[|y - t|] \le \frac{\rho}{\epsilon(1-\gamma)}\left(1 - \gamma[1 + \ln(\frac{1}{\gamma})]\right)$. This completes the proof of Lemma 2. $\square$

Theorem 1. *For any $\theta \in [0, 1]$ and $\gamma \ge 0$, $\mathcal{M}_{\text{rob}}$ is $(\theta, \rho, \epsilon, \delta)$-robust and $(0, \beta, \gamma)$-accurate, where $\delta = 2e^\epsilon\theta$ and $\beta = \frac{\rho}{\epsilon(1-\gamma)}\left(1 - \gamma[1 + \ln(\frac{1}{\gamma})]\right)$.*

Proof. Using Lemma 1 and Lemma 2, the theorem trivially holds. $\square$

3.2 Adding Robustness to a Mechanism

Consider a randomized mechanism $\mathcal{M} : A \to B$ which has some privacy and accuracy bounds and a randomized mechanism $\mathcal{RM} : B \to B$ which is robust and has some accuracy bounds. We want to construct a new mechanism which is robust, at least as private as $\mathcal{M}$, and doesn't compromise much on accuracy as compared to M. Let $x \in A$ be the input. Consider the following mechanism:

Mechanism 2. *Run $\mathcal{M}$ on x to get an output $y \in B$ then run $\mathcal{M}_{\text{rob}}$ on y to get $z \in B$, output z.*

Theorem 2. *Let $\rho, \theta, \epsilon, \delta$ be non-negative real numbers with $\delta, \theta \le 1$. Then Mechanism 2 achieves* (i) *$(\theta, \rho, \epsilon, \delta)$-robustness, where $\delta = 2e^\epsilon\theta$,* (ii) *$(\epsilon_p, \delta_p)$-differential privacy, if $\mathcal{M}$ is (ϵ_p, δ_p)-differentially private, and* (iii) *$(\alpha_p, \beta_p, \gamma_p)$-accuracy, if M is (α, β, γ)-accurate, where $\alpha_p = \alpha, \beta_p = \beta + \beta', \gamma_p = \gamma + \gamma'$ and β', γ' are such that $\gamma' \ge 0$ is arbitrary and $\beta' = \frac{\rho}{\epsilon(1-\gamma')}\left(1 - \gamma'[1 + \ln(\frac{1}{\gamma'})]\right)$.*

4 Differential Privacy for Randomized Queries

Suppose the database is $\mathbf{x} \in \mathcal{X}^n$ and $\theta \in [0, 1], \epsilon > 0$ be fixed constants. Consider the following randomized mechanism for a randomized query $f : \mathcal{X}^n \to \mathbb{R}$:

Mechanism 3. *Sample y and z independently and according to the probability distributions $f(\mathbf{x})$ and $Lap(\frac{S^\theta(f)}{\epsilon})$, respectively, and output $y + z$.*

Theorem 3. *For the fixed constants $\theta' \in [0, 1], \epsilon' > 0, \gamma \ge 0$, Mechanism 3 achieves $(\theta', S^\theta(f), \epsilon', \delta')$-robustness, (ϵ, δ)-differential privacy and $(0, \beta, \gamma)$-accuracy, where $\delta' = 2e^{\epsilon'}\theta'$, $\delta = 2e^\epsilon\theta$ and $\beta = \frac{S^\theta(f)}{\epsilon(1-\gamma)}\left(1 - \gamma[1 + \ln(\frac{1}{\gamma})]\right)$.*

Proof. Observe that f can be treated as a mechanism for f with (0,0,0)-accuracy and $(\infty, 0)$-privacy and that Mechanism 3 is equivalent to $\mathcal{M}_{\text{rand}} := \mathcal{M}_{\text{rob}} \circ f$, where in $\mathcal{M}_{\text{rob}}$, the Laplace noise has the parameter $\rho = S^\theta(f)$. The robustness and accuracy bounds are obtained from Theorem 2. To show that $\mathcal{M}_{\text{rand}}$ is (ϵ, δ)-differentially private, consider any two neighbouring databases $\mathbf{x}, \mathbf{x}' \in \mathcal{X}^n$. Since $\mathbf{x} \sim \mathbf{x}'$, it follows from Definition 6 that $W_\infty^\theta(f(\mathbf{x}), f(\mathbf{x}')) \le S^\theta(f)$. This, together with the fact that Mechanism 3 is $(\theta', S^\theta(f), \epsilon', \delta')$-robust, implies that $Pr[\mathcal{M}_{\text{rand}}(\mathbf{x}) \in S] \le e^\epsilon Pr[\mathcal{M}_{\text{rand}}(\mathbf{x}') \in S] + \delta$, where $\delta = (e^\epsilon + 1)2\theta$. But the definition of robustness requires that $W_\infty^\theta(\mathcal{M}_{\text{rand}}(\mathbf{x}), \mathcal{M}_{\text{rand}}(\mathbf{x}')) \le S^\theta(f)$, not $W_\infty^\theta(f(\mathbf{x}), f(\mathbf{x}')) \le S^\theta(f)$. How do we fix this? Note that Claim 1 is for W^θ, not W_∞^θ! $\square$

5 Histogram Mechanism

In this section, we propose a new mechanism to get a histogram from a given (correct) histogram in a differentially private manner.

Mechanism 4. *For each bar of the histogram, independently add a noise according to the following probability distribution:*

$$\pi_q(x) = \begin{cases} \frac{1}{1-e^{-\frac{\sqrt{q}}{2}}} Lap(x \mid -\frac{q}{2}, \sqrt{q}) & -q < x < 0 \\ 0 & elsewhere \end{cases}$$

and then round each of the bar to the nearest integer. If any of the bar becomes negative, then set it to 0. Output the resulting histogram.

To prove that the above mechanism is (ϵ, δ)-differentially private, first we show below in Lemma 5 that, if in Mechanism 4, we output the histogram before rounding and setting negative values to 0, then we get (ϵ, δ)-differential privacy.

Lemma 3. *For any $q \geq 1$, if in Mechanism 4 we output the histogram before rounding and setting negative values to 0, then we get (ϵ, δ)-differential privacy, where $\epsilon = \frac{1}{\sqrt{q}}$ and $\delta = \frac{4}{\sqrt{q}} e^{-\frac{\sqrt{q}}{2}}$.*

Lemma 5 is proved in Appendix C. Observe that rounding and setting negative values to 0 is a post-processing step. Since Lemma 5 is (ϵ, δ)-differentially private and post-processing preserves differential privacy [1, Proposition 2.1], it follows that Mechanism 4 is also (ϵ, δ) private. The following theorem establishes $(qt, 0, 0)$-accuracy of Mechanism 4.

Theorem 4. *Mechanism 4 achieves $(\alpha, 0, 0)$ accuracy with $\alpha = qt$.*

6 Using the Histogram Mechanism for General Statistical Queries

We now show how histogram mechanism can be used for answering a general query. For a query which desires some statistic of the database, we construct an appropriate function, $f^{'}$ which takes the histogram given by the above mechanism and outputs the desired statistic within some error. The input database is fist bucketed, i.e, elements are rounded off to a certain number of elements and modified according to the above mentioned histogram mechanism. The modified histogram given to the above function. The output thus obtained is guaranteed to be accurate and private. To find the accuracy and private parameters we prove the following theorems.

Theorem 5. *Let $M_1 : A \to B$ be a mechanism which is $(\alpha_1, \beta_1, \gamma_1)$ accurate for a function $f_1 : A \to B$ and let $M_2 : B \to C$ be a mechanism which is $(\alpha_2, \beta_2, \gamma_2)$ accurate for a function $f_2 : B \to C$. Let d_1, d_2 be distortion functions on A, B, respectively. Then the composite mechanism $M_2 \circ M_1 : A \to C$ is (α, β, γ)-accurate for the function $f_2 \circ f_1$, where $\alpha = \alpha_1 + \sigma_{f_1}(\alpha_2), \beta = \beta_2 + \tau_{M_2}^{\gamma_1}(\beta_1), \gamma = \gamma_1 + \gamma_2$.*

Proof. We prove the theorem using a hybrid argument. Consider the hybrid mechanism $M_2 \circ f_1$, along with $f_2 \circ f_1$ and $M_2 \circ M_1$. Below, we compute the Wasserstein distances between $M_2 \circ M_1$ & $M_2 \circ f_1$ and between $M_2 \circ f_1$ & $f_2 \circ f_1$. Then we use the triangle inequality from Claim 2 to find the Wasserstein distance between $M_2 \circ M_1$ and $f_2 \circ f_1$.

For a given database $\mathbf{x}$, since M_1 is $(\alpha_1, \beta_1, \gamma_1)$-accurate mechanism for f_1, there exists a random variable X' which is a distribution over the α_1-distorted databases from $\mathbf{x}$, such that

$$W^{\gamma_1}(f_1(X'), M_1(\mathbf{x})) \leq \beta_1. \tag{12}$$

Now, applying the mechanism M_2 over the distributions $M_1(\mathbf{x})$, $f_1(X')$ may increase the error by at most $\tau_{M_2}^{\gamma_1}(\beta_1)$ (see Definition 11), which gives

$$W^{\gamma_1}(M_2(f_1(X')), M_2(M_1(\mathbf{x}))) \leq \tau_{M_2}^{\gamma_1}(\beta_1). \tag{13}$$

Now, we bound the Wasserstein distance between $M_2 \circ f_1$ and $f_2 \circ f_1$. For any database $\mathbf{x}_1$ drawn from X', since M_2 is an $(\alpha_2, \beta_2, \gamma_2)$-accurate mechanism for f_2, there exists a random variable X'', which is a distribution over α_2-distorted databases from $f_1(\mathbf{x}_1)$, such that $W^{\gamma_2}(f_2(X''), M_2(f_1(\mathbf{x}_1))) \leq \beta_2$. Since this holds for every $\mathbf{x}_1$ in the support of X' (note that X'' may depend on $\mathbf{x}_1$), we have

$$W^{\gamma_2}(f_2(X''), M_2(f_1(X'))) \leq \beta_2. \tag{14}$$

Combining (14) and (14) and using Claim 2, we get

$$W^{\gamma_1+\gamma_2}(f_2(X''), M_2(M_1(\mathbf{x}))) \leq \beta_2 + \tau^{\gamma_1}_{M_2}(\beta_1). \tag{15}$$

Note that X'' is a distribution over B and we need to find a distribution over the $\alpha = (\alpha_1 + \sigma_{f_1}(\alpha_2))$-distorted databases from the database $\mathbf{x}$. For this $\mathbf{x} \in A$, consider the corresponding distribution of X' guaranteed by the mechanism M_1, which is defined over all the elements $\mathbf{x}_1 \in A$ such that $d_1(\mathbf{x}, \mathbf{x}_1) \leq \alpha_1$. For every $\mathbf{x}_1$ drawn from X', consider the distribution of X'' guaranteed by M_2, which is defined over all the elements in $\mathbf{x}_1' \in B$ such that $d_2(f_1(\mathbf{x}_1), \mathbf{x}_1') \leq \alpha_2$.

Since $d_2(f_1(\mathbf{x}_1), \mathbf{x}_1') \leq \alpha_2$, it follows from Definition 10 that there exists a database $\mathbf{x}_2 \in A$, such that $f_1(\mathbf{x}_2) = \mathbf{x}_1'$ and that $d_1(\mathbf{x}_1, \mathbf{x}_2) \leq \sigma_{f_1}(\alpha_2)$, Let X_2 denote the distribution over all such $\mathbf{x}_2$'s in A. It follows from the triangular inequality of d_1 that $d_1(\mathbf{x}, \mathbf{x}_2) \leq \alpha_1 + \sigma_{f_1}(\alpha_2)$. Thus, X_2 is a distribution over databases in A, which are at most $(\alpha_1 + \sigma_{\alpha, f_1}(\alpha_2))$-distorted from $\mathbf{x}$ and satisfy

$$W^{\gamma_1+\gamma_2}(f_2(f_1(X_2)), M_2(M_1(\mathbf{x}))) \leq \beta_2 + \tau^{\gamma_1}_{M_2}(\beta_1). \tag{16}$$

This implies that the composite mechanism $M_2 \circ M_1$ is (α, β, γ)-accurate, where $\alpha = \alpha_1 + \sigma_{f_1}(\alpha_2), \beta = \beta_1 + \tau^{\gamma_1}_{M_2}(\beta_1), \gamma = \gamma_1 + \gamma_2$. This completes the proof of Theorem 5. □

Now, we present another composition theorem, which we prove in Appendix D.

Theorem 6. *Let $M_1 : A \to B$ be a neighbourhood preserving mechanism for a function $f_1 : A \to B$, and let $M_2 : B \to C$ be a mechanism which is (ϵ, δ)-differential private for a function $f_2 : B \to C$. Then the composite mechanism, $M_2 \circ M_1 : A \to C$ is (ϵ, δ)-differential private.*

Now, we show how appropriate mechanisms can be composed to achieve good accuracy and privacy guarantees for computing a statistic on a database. We consider databases which consists of positive real number numbers. The bucketing is done with t number of buckets, i.e, rounding each element of database to the nearest multiple of B/t.

Lemma 4. *The bucketing followed by Mechanism 4 is an $(\alpha_p, \beta_p, 0)$-accurate and (ϵ_p, δ_p)-differentially private mechanism for the identity function, with $\alpha_p = qt, \beta_p = \frac{B}{2t}, \epsilon_p = \frac{1}{\sqrt{q}}, \delta_p = \frac{4}{\sqrt{q}} e^{-\frac{\sqrt{q}}{2}}$.*

6.1 Case Study

We now show the application of the histogram for computing Max and support of a database in an accurate and private manner.

6.1.1 Maximum

Let d be a distance metric on $\mathcal{X}$. We define the function $f : \mathcal{N}^{\mathcal{X}} -> \mathcal{X}$, as one which takes a database as input and outputs the largest element. Lets denote the distance metric over input space by d_1 and on output space by d_2. For any two databases x_1 and x_2, $d_1(x_1, x_2) =$ the maximum d-distance each element of x_1 is to be modified to get x_2 or vice-versa (both are equivalent) and $d_2(x_1, x_2) =$ d-distance between x_1 and x_2 in $\mathcal{X}$ space. Note that f is a perfectly accurate mechanism for maximum. Also, if for any two databases x_1 and x_2, $d_1(x_1, x_2) = l$ then $d_2(f(x_1), f(x_2)) = l$ because we don't need to modify the max of x_1 by more than l to get max of x_2. Using this we get that the error-sensitivity function for f is identity, i.e, $\tau_f(\beta') = \beta'$. Because if two distribution over databases have d_1-wasserstein distance β, then the output distribution will have d_2-wasserstein β.

Now, using Theorem 5 and Theorem 6, we have that, the histogram mechanism for finding max is $(\alpha_p, \beta_p, 0)$ accurate and (ϵ_p, δ_p) private where $\alpha_p, \beta_p, \epsilon_p$ and δ_p are defined in Lemma 4.

6.1.2 Support

Here we define the function $f : \mathcal{N}^{\mathcal{X}} - > \mathcal{N}^{\mathcal{X}}$, which takes a database as input and outputs another database. The function just removes duplicate entries from the input database and outputs the resulting database. Let d be a metric over X. The distance metrics d_1, d_2 are defined as follows for any two databases x_1 and x_2: $d_1(x_1, x_2) = d_2(x_1, x_2)$ = the minimum distance l (according to d) such that we can move each element of x_1 and x_2 by atmost l and ensure that after moving there is no element which is in only x_1 or only x_2. In other words, after moving, the set of elements in x_1 must be equal to the set of elements in x_2. Again, f is a perfectly accurate mechanism for support. Also the error-sensitivity function for f is identity, i.e, Also, if for any two databases x_1 and x_2, $d_1(x_1, x_2) = l$ then $d_2(f(x_1), f(x_2)) = l$ because if we could move each element of input databases by a distance β' and satisfy the condition then the output is just the same entries just with duplicates removed so we can use the same distance to move them to satisfy the constraints. Now, $\tau_f(\beta') = \beta'$ using the same argument as in the case of maximum. Now, using Theorem 5 and Theorem 6, we have that, the histogram mechanism for finding max is $(\alpha_p, \beta_p, 0)$ accurate and (ϵ_p, δ_p) private where $\alpha_p, \beta_p, \epsilon_p$ and δ_p are defined in Lemma 4.

References

[1] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407, August 2014. 3, 8, 16, 18, 20

A Omitted Details from Section 2

Claim 3.

$$\inf_{\phi \in \Phi^\theta(p,q)} \max_{(x,y)\leftarrow \phi} |x-y| = \inf_{\substack{p',q':\\ \Delta(p,p')\le\theta,\\ \Delta(q,q')\le\theta}} W_\infty(p',q'),$$

where p' and q' are defined over the same alphabets where p and q are defined, respectively.

Proof. fill in. □

We don't need the above claim, right?

Claim (Restating Claim 1)**.** *let X and Y be a random variable with distributions μ_1, μ_2. Z be a an independent noise random variable with a distribution μ_3. Then we have,*

$$W^\gamma(X,Y) = W^\gamma(X+Z, Y+Z)$$

i.e, convolution of distributions with the same independent noise doesn't change wasserstein distance.

Proof. First, we show that applying convolution of distributions with another doesn't increase the wasserstein distance between them.

Let p, q, r be three distributions, with wasserstein distance between p,q as β and π be the optimal transfer which achieves this . Let I_1, I_2 be two distributions obtained by convolving p, q with r respectively. i.e,

$$Pr(I_1 = x) = \int_Z p(x-z)r(z)$$
$$Pr(B+C = y) = \int_Z q(y-z)r(z)$$

From the above equations, we can construct a transfer from I_1, I_2 as follows. Let z be drawn from r with a probability $p(z)$. In this event, we transfer I_1 to I_2 using the policy, π + z. This policy also transfers wasserstein distance of β for all z. Hence, it's expectation over z is also β. Since, wasserstein distance is an infimum of all transfers, we have, the wasserstein distance between I_1, I_2 to be at most β. Hence, proved.

Now, for the random variables X, Y, Z. The distributions of X + Z, Y + Z are given by the following convolutions.

$$Pr(A+C = x) = \int_Z \mu_1(x-z)\mu_3(z)$$
$$Pr(B+C = y) = \int_Z \mu_2(y-z)\mu_3(z)$$

From the above reasoning, we conclude that the wasserstein distance between X + Z, Y + Z is atmost that of X, Y.

Note that the distributions of X and Y can be obtained by inverse convolution of Z with that of X + Z, Y + Z. This is equivalent to convolution with an appropriate distribution Z'. This is true for discrete variables, convolution is matrix multiplication with a kernel, so inverse convolution is just multiplying with inverse of the kernel. How does it go for continuous variables ? . Again, as convolving with a distribution doesn't increase the wasserstein distance, we conclude the wasserstein distance between X + Z, ,Y + Z is atmost that of X, Y.

Combining both these results, we have that, wasserstein distance between X, Y is same as wasserstein distance between X + Z, Y + Z. □

Claim (Restating Claim 2)**.** *θ-Wasserstein distance follows the below triangular inequality. For distributions, p, q and r, we have,*

$$W^{\gamma_1+\gamma_2}(p,r) \le W^{\gamma_1}(p,q) + W^{\gamma_2}(q,r)$$

Proof. fill in. □

Let $\mathcal{X}$ be a finite set, and let $\mathbf{x} := (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ represent a database with n elements, where $x_i \in \mathcal{X}$ is the i-th element. Let $f : \mathcal{X}^n \to \mathcal{Y}$ be a randomized function, where $\mathcal{Y} = \{y_1, y_2, \ldots, y_k\}$ is a discrete set for some finite $k \in \mathbb{N}$. For every $\mathbf{x} \in \mathcal{X}^n$, we have that $\sum_{i=1}^{k} \Pr[f(\mathbf{x}) = y_i] = 1$.

B Omitted Details from Section 3

B.1 Robust Mechanism for Identity Function over $\mathbb{R}$

We will now show a mechanism $\mathcal{M}_{\text{rob}} : \mathbb{R} \to \mathbb{R}$ which is a robust mechanism for identity function, i.e., the output corresponding to any input is the same as input. Now Suppose the input distribution is x. Let ρ be any non-negative real number and ϵ be any positive real number. Consider the following mechanism for the identity function:

> **Mechanism** (Restating Mechanism 1)**.** $\mathcal{M}_{\text{rob}}$*: On input y, sample z according to the probability distributions Lap$(\frac{\rho}{\epsilon})$ and output $y + z$.*

Lemma (Restating Lemma 1)**.** *For a fixed constant $\theta \in [0, 1]$, $\mathcal{M}_{\text{rob}}$ achieves $(\theta, \rho, \epsilon, \delta)$-robustness where $\delta = 2e^{\epsilon}\theta$.*

Proof. Fix a constant $\theta \geq 0$. Suppose q and q' denote two input distributions, and let p and p' denote the corresponding output probability distributions of $\mathcal{M}_{\text{rob}}$. Suppose $W_{\infty}^{\theta}(p, p') \leq \rho$. Then by Claim 1, $W_{\infty}^{\theta}(q, q') \leq \rho$. Let $b := \frac{\rho}{\epsilon}$ and $Z \sim Lap(b)$. We need to show that $\forall S \subseteq \mathbb{R}$, $\Pr_{y \leftarrow p}[y \in S] \leq e^{\epsilon} \Pr_{y \leftarrow p'}[y \in S] + \delta$.

$$\Pr_{y \leftarrow p}[y \in S] = \int_{t \in S} p(t) \cdot dt \tag{17}$$

$$= \int_{t \in S} \big[\int_{-\infty}^{\infty} q(s) \cdot p_Z(t - s) \cdot ds \big] \cdot dt \tag{18}$$

Let $\phi^{\theta} \in \Phi^{\theta}(q, q')$ be a joint distribution. Let M^{ϕ} denote $\max_{(x,y) \leftarrow \phi^{\theta}} |x - y|$ for the distribution ϕ^{θ}. For $i \in \mathbb{R}$, let $\theta_1(i) := |\phi_1^{\theta}(i) - q(i)|$ and $\theta_2(i) := |\phi_2^{\theta}(i) - q'(i)|$. Note that, by definition of the distribution ϕ^{θ}, we have $\Delta(\phi_1^{\theta}, q) + \Delta(\phi_2^{\theta}, q') \leq \theta$ which implies $\int_{-\infty}^{\infty} \theta_1(i) \cdot di + \int_{-\infty}^{\infty} \theta_2(i) \cdot di \leq \theta$. Using these in (18) we get the following:

$$\begin{aligned}
\Pr_{y \leftarrow p}[y \in S] &\leq \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} (\phi_1^{\theta}(i) + \theta_1(i)) \cdot e^{\frac{-|t-i|}{b}} \cdot di \big] \cdot dt \\
&= \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} \phi_1^{\theta}(i) \cdot e^{\frac{-|t-i|}{b}} \cdot di \big] \cdot dt + \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} \theta_1(i) \cdot e^{\frac{-|t-i|}{b}} \cdot di \big] \cdot dt \\
&= \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} \phi_1^{\theta}(i) \cdot e^{\frac{-|t-i|}{b}} \cdot di \big] \cdot dt + \int_{-\infty}^{\infty} \theta_1(i) \cdot \underbrace{\int_{t \in S} \frac{1}{2b} e^{\frac{-|t-i|}{b}} \cdot dt}_{\leq 1} \cdot di
\end{aligned} \tag{19}$$

By properties of joint distributions, we have $\phi_1^{\theta}(i) = \int_{-\infty}^{\infty} \phi^{\theta}(i, j) \cdot dj$ and $\phi_2^{\theta}(j) = \int_{-\infty}^{\infty} \phi^{\theta}(i, j) \cdot di$; and by triangle inequality we have $|t - i| \geq |t - j| - |i - j|$. Let $\int_{-\infty}^{\infty} \theta_l(i) \cdot di$ be ω_l, $l \in \{1, 2\}$. Substituting all these in (19) we get the following:

$$\Pr_{y \leftarrow p}[y \in S] \leq \int_{t \in S} \big[\frac{1}{2b} \int_{-\infty}^{\infty} [\int_{-\infty}^{\infty} \phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|+|i-j|}{b}} \cdot dj] \cdot di \big] \cdot dt + \omega_1. \tag{20}$$

Observe that $\phi^{\theta}(i, j)$ is non-zero only if $|i - j| \leq M^{\phi}$ (by definition of M^{ϕ}). Using this for every $i, j \in \mathbb{R}$, the integrand in (20) can be upper-bounded as follows:

$$\phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|}{b}} e^{\frac{|i-j|}{b}} \leq \phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|}{b}} e^{\frac{M^{\phi}}{b}} = \phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|}{b}} e^{\frac{M^{\phi}}{b}}.$$

Substituting this in (20) gives

$$\Pr_{y \leftarrow p}[y \in S] \leq \int_{t \in S} \big[\frac{e^{\frac{M^{\phi}}{b}}}{2b} \int_{-\infty}^{\infty} [\int_{-\infty}^{\infty} \phi^{\theta}(i, j) \cdot e^{\frac{-|t-j|}{b}} \cdot dj] \cdot di \big] \cdot dt + \omega_1$$

$$
\begin{aligned}
&= \int_{t\in S} \Big[\frac{e^{\frac{M^\phi}{b}}}{2b} \int_{-\infty}^{\infty} \Big[\int_{-\infty}^{\infty} \phi^\theta(i,j) \cdot e^{\frac{-|t-j|}{b}} \cdot di\Big] \cdot dj\Big] \cdot dt + \omega_1 \\
&= \int_{t\in S} \Big[\frac{e^{\frac{M^\phi}{b}}}{2b} \int_{-\infty}^{\infty} \phi_2^\theta(j) \cdot e^{\frac{-|t-j|}{b}} \cdot dj\Big] \cdot dt + \omega_1 \\
&\le \int_{t\in S} \Big[\frac{e^{\frac{M^\phi}{b}}}{2b} \int_{-\infty}^{\infty} (q'_j + \theta_2(j)) \cdot e^{\frac{-|t-j|}{b}} \cdot dj\Big] \cdot dt + \omega_1 \quad (\text{since } \theta_2(j) = |\phi_2^\theta(j) - q'(j)|) \\
&\le \int_{t\in S} \Big[\frac{e^{\frac{M^\phi}{b}}}{2b} \int_{-\infty}^{\infty} q'_j \cdot e^{\frac{-|t-j|}{b}} \cdot dj\Big] \cdot dt + \int_{t\in S} \Big[\frac{e^{\frac{M^\phi}{b}}}{2b} \int_{-\infty}^{\infty} \theta_2(j) \cdot e^{\frac{-|t-j|}{b}} \cdot dj\Big] \cdot dt + \omega_1 \\
&\le \int_{t\in S} \Big[\frac{e^{\frac{M^\phi}{b}}}{2b} \int_{-\infty}^{\infty} q'_j \cdot e^{\frac{-|t-j|}{b}} \cdot dj\Big] \cdot dt + e^{\frac{M^\phi}{b}} \int_{-\infty}^{\infty} \theta_2(j) \underbrace{\Big[\int_{t\in S} \frac{1}{2b} \cdot e^{\frac{-|t-j|}{b}} \cdot dt\Big]}_{\le 1} \cdot dj + \omega_1 \\
&\le \int_{t\in S} \Big[\frac{e^{\frac{M^\phi}{b}}}{2b} \int_{-\infty}^{\infty} q'_j \cdot e^{\frac{-|t-j|}{b}} \cdot dj\Big] \cdot dt + e^{\frac{M^\phi}{b}} \cdot \omega_2 + \omega_1 \\
&= \int_{t\in S} \Big[e^{\frac{M^\phi}{b}} \int_{-\infty}^{\infty} p'(t) \cdot dj\Big] \cdot dt + e^{\frac{M^\phi}{b}} \cdot \omega_2 + \omega_1 \\
&= e^{\frac{M^\phi}{b}} \Pr_{y\leftarrow p'}[y \in S] + e^{\frac{M^\phi}{b}} \cdot \omega_2 + \omega_1
\end{aligned}
$$

Now, by the limiting case of ϕ^θ to be optimal, M^ϕ tends to ρ, thus we have,

$$
\begin{aligned}
\Pr_{y\leftarrow p}[y \in S] &\le e^{\frac{\rho}{b}} \Pr_{y\leftarrow p'}[y \in S] + e^{\frac{\rho}{b}} \cdot \omega_2 + \omega_1 \\
&\le e^{\frac{\rho}{b}} \Pr_{y\leftarrow p'}[y \in S] + e^{\frac{\rho}{b}} \cdot 2\theta \quad (\text{ since } \omega_1 + \omega_2 \le 2\theta, \text{ RHS is maximized at } \omega_2 = 2\theta) \\
&= e^{\epsilon} \Pr_{y\leftarrow p'}[y \in S] + \delta, \qquad \text{where } \delta = 2\theta e^{\epsilon}
\end{aligned}
$$

□

Lemma (Restating Lemma 2)**.** *For every constant $\gamma' \ge 0$, $\mathcal{M}_{\text{rob}}$ is $(0, \beta', \gamma')$-accurate, where $\beta' = \frac{\rho}{\epsilon(1-\gamma')}\Big(1 - \gamma'[1 + \ln(\frac{1}{\gamma'})]\Big)$. Note that if $\gamma' = 0$, $\beta' = \frac{\rho}{\epsilon}$.*

Proof. Fix a constant $\gamma \ge 0$ and take an arbitrary distribution (over $\mathbb{R}$) and lets denote it by x. Let q denote the output distribution of $\mathcal{M}_{\text{rob}}$ when the input is drawn from x. Let b denote $\frac{\rho}{\epsilon}$. We want to show that $W^\gamma(x, q) \le \beta$, for the above-mentioned β. By definition of W^γ from (2) we have $W^\gamma(x, q) = \inf_{\phi\in\Phi^\gamma(x,q)} \mathbb{E}_{(y,t)\leftarrow\phi}[|y - t|]$.

Consider the following ϕ^*:

$$
\phi^*(i,t) = \begin{cases} 0 & \text{if } t < -b\ln(\frac{1}{\gamma}) + i \text{ or } t > b\ln(\frac{1}{\gamma}) + i; \\ \frac{1}{1-\gamma} Lap(t|b,i)x(i) & \text{if } t \in [-b\ln(\frac{1}{\gamma}) + i, b\ln(\frac{1}{\gamma}) + i]. \end{cases}
$$

It can be verified that $\Delta(\phi_1^*, x) = 0$ and $\Delta(\phi_2^*, q) \le \gamma$, which implies that $\phi^* \in \Phi^\gamma(x, q)$. This in turn implies that $W^\gamma(x, q) \le \mathbb{E}_{(y,t)\leftarrow\phi^*}[|y - t|]$. We show below that $\mathbb{E}_{(y,t)\leftarrow\phi^*}[|y - t|] \le \frac{b}{(1-\gamma)}\Big(1 - \gamma[1 + \ln(\frac{1}{\gamma})]\Big)$. This will prove Lemma 2.

$$
\begin{aligned}
\mathbb{E}_{(y,t)\leftarrow\phi^*}[|y - t|] &= \int_{w=-\infty}^{w=\infty} |w| \cdot \Pr_{(y,t)\leftarrow\phi^*}[|y - t| = w] \cdot dw \\
&= \int_{w=0}^{w=\infty} w \cdot \Pr_{(y,t)\leftarrow\phi^*}[|y - t| = w] \cdot dw \\
&= \int_{w=0}^{w=\infty} w \cdot \Big[\Pr_{(y,t)\leftarrow\phi^*}[y - t = w] + \Pr_{(y,t)\leftarrow\phi^*}[y - t = -w]\Big] \cdot dw
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{w=0}^{w=\infty} w \cdot \big[\int_{-\infty}^{\infty} \phi^*(i, i-w) \cdot di + \int_{-\infty}^{\infty} \phi^*(i, i+w) \cdot di \big] \cdot dw \\
&= \int_{w=0}^{w=\infty} w \cdot \big[\int_{-\infty}^{\infty} 2\phi^*(i, i+w) \cdot di \big] \cdot dw \\
& \qquad\qquad\qquad\qquad (\text{Since } \phi^*(i,i-w) = \phi^*(i,i+w)) \\
&= \int_{-\infty}^{\infty} \big[\int_{0}^{\infty} w \cdot 2\phi^*(i, i+w) \cdot dw \big] \cdot di
\end{aligned}
$$

Let $d = b\ln(\frac{1}{\gamma})$. By definition of ϕ^*, $\phi^*(i, i+w)$ is non-zero only if $i + w \in [i-d, i+d]$, which implies that $w \in [-d, d]$. Using this in above gives:

$$
\begin{aligned}
\mathbb{E}_{(y,t)\leftarrow\phi^*}[|y-t|] &= \int_{-\infty}^{\infty} \big[\int_{0}^{d} w \cdot 2\phi^*(i, i+w) \cdot dw \big] \cdot di \\
&= \int_{-\infty}^{\infty} x_i \big[\int_{0}^{d} w \cdot \frac{1}{(1-\gamma)} \left(\frac{1}{b} e^{\frac{-|i+w-i|}{b}} \right) \cdot dw \big] \cdot di \\
&= \frac{1}{b(1-\gamma)} \int_{-\infty}^{\infty} x(i) \big[\int_{0}^{d} w \cdot \left(e^{\frac{-|w|}{b}} \right) \cdot dw \big] \cdot di \\
&= \frac{1}{b(1-\gamma)} \int_{-\infty}^{\infty} x(i) \big[\int_{0}^{d} w \cdot \left(e^{\frac{-w}{b}} \right) \cdot dw \big] \cdot di \\
&= \frac{1}{b(1-\gamma)} \int_{-\infty}^{\infty} x(i) \big[\int_{0}^{d/b} (bw) \cdot e^{-w} \cdot (b \cdot dw) \big] \cdot di \\
&= \frac{b}{(1-\gamma)} \int_{-\infty}^{\infty} x(i) \big[\int_{0}^{\ln(\frac{1}{\gamma})} w \cdot e^{-w} \cdot dw \big] \cdot di \qquad (\text{since } d = b\ln(\tfrac{1}{\gamma})) \\
&= \frac{b}{(1-\gamma)} \int_{-\infty}^{\infty} x(i) \big[(-we^{-w} - e^{-w})\big|_{w=0}^{w=\ln(\frac{1}{\gamma})} \big] \cdot di \\
&= \frac{b}{(1-\gamma)} \int_{-\infty}^{\infty} x(i) \big(1 - \gamma[1 + \ln(\frac{1}{\gamma})] \big) \cdot di \\
&= \frac{b}{(1-\gamma)} \big(1 - \gamma[1 + \ln(\frac{1}{\gamma})] \big) \\
&= \frac{\rho}{\epsilon(1-\gamma)} \big(1 - \gamma[1 + \ln(\frac{1}{\gamma})] \big) \qquad (\text{since } b = \tfrac{\rho}{\epsilon})
\end{aligned}
$$

Note that at $\gamma = 0$, we have $\mathbb{E}_{(y,t)\leftarrow\phi^*}[|y-t|] = \frac{\rho}{\epsilon}$. □

Theorem 7. *For any $\theta \in [0,1]$ and $\gamma \geq 0$, $\mathcal{M}_{\text{rob}}$ is $(\theta, \rho, \epsilon, \delta)$-robust and $(0, \beta, \gamma)$-accurate, where $\delta = 2e^\epsilon\theta$ and $\beta = \frac{\rho}{\epsilon(1-\gamma)} \left(1 - \gamma[1 + \ln(\frac{1}{\gamma})] \right)$.*

Proof. Using Lemma 1 and Lemma 2, the theorem trivially holds. □

B.2 Adding Robustness to a Mechanism

Consider a randomized mechanism $\mathcal{M} : A \to B$ which has some privacy and accuracy bounds and a randomized mechanism $\mathcal{M}_{\text{rob}} : B \to B$ which is robust and has some accuracy bounds. We want to construct a new mechanism which is robust, atleast as private as $\mathcal{M}$ and doesn't compromise much on accuracy as compared to $\mathcal{M}$. Let $x \in A$ be the input. Consider the following mechanism:

Mechanism (Restating Mechanism 2)**.** *Run $\mathcal{M}$ on x to get an output $y \in B$ then run $\mathcal{M}_{\text{rob}}$ on y to get $z \in B$, output z.*

Theorem (Restating Theorem 2)**.** *Let $\rho, \theta, \epsilon, \delta$ be non-negative real numbers with $\delta, \theta \leq 1$. Then Mechanism 2 achieves* (i) *$(\theta, \rho, \epsilon, \delta)$-robustness, where $\delta = 2e^{\epsilon}\theta$,* (ii) *$(\epsilon_p, \delta_p)$-differential privacy, if $\mathcal{M}$ is (ϵ_p, δ_p)-differentially private, and* (iii) *$(\alpha_p, \beta_p, \gamma_p)$-accuracy, if M is (α, β, γ)-accurate, where $\alpha_p = \alpha, \beta_p = \beta + \beta', \gamma_p = \gamma + \gamma'$ and β', γ' are such that $\gamma' \geq 0$ is arbitrary and $\beta' = \frac{\rho}{\epsilon(1-\gamma')}\left(1 - \gamma'[1 + \ln(\frac{1}{\gamma'})]\right)$.*

Proof. Observe that Mechanism 2 is equivalent to $\mathcal{M}_{\text{rob}} \circ \mathcal{M}$. First we prove the accuracy guarantee. From Lemma 2, we have that $\mathcal{M}_{\text{rob}}$ is $(0, \beta', \gamma')$ accurate, where $\gamma' \geq 0$ is arbitrary and $\beta' = \frac{\rho}{\epsilon(1-\gamma')}\left(1 - \gamma'[1 + \ln(\frac{1}{\gamma'})]\right)$. Now, let us compute the error-sensitivity function of $\mathcal{M}_{\text{rob}}$.

$$\begin{aligned}
\tau^{\theta}_{\mathcal{M}_{\text{rob}}}(\beta) &= \max_{\substack{x \sim p, x' \sim q: \\ W^{\theta}(p,q) \leq \beta}} W^{\theta}(\mathcal{M}_{\text{rob}}(x), \mathcal{M}_{\text{rob}}(x')) \\
&\overset{\text{(a)}}{=} \max_{\substack{x \sim p, x' \sim q: \\ W^{\theta}(p,q) \leq \beta}} W^{\theta}(p, q) \\
&= \beta.
\end{aligned}$$

Here (a) follows from Claim 1 and that $\mathcal{M}_{\text{rob}}$ adds independent laplacian noise. Thus, we have that the required error sensitivity function is identity. Now, if $\mathcal{M}$ is (α, β, γ)-accurate, then by Theorem 5, we have that $\mathcal{M}_{\text{rob}} \circ \mathcal{M}$ is $(\alpha_p, \beta_p, \gamma_p)$-accurate with $\alpha_p = \alpha, \beta_p = \beta' + \beta, \gamma_p = \gamma' + \gamma$.

For the privacy guarantee, since $\mathcal{M}$ is (ϵ_p, δ_p)-differentially private and post-processing preserves differential privacy [1, Proposition 2.1], $\mathcal{M}_{\text{rob}} \circ \mathcal{M}$ is also (ϵ_p, δ_p)-differentially private.

The proof of robustness is straightforward. By Definition 9, if *any* two input distributions gives close (in terms of the Wasserstein distance) output distributions, then they are also close in terms of differential privacy sense. Since the requirement is only on the output distributions, it follows that if a mechanism A is robust, then $A \circ B$ is also robust for every mechanism B, with exactly the same parameters. This implies that, since $\mathcal{M}_{\text{rob}}$ is $(\theta, \rho, \epsilon, 2e^{\epsilon}\theta)$-robust (see Lemma 1), $\mathcal{M}_{\text{rob}} \circ \mathcal{M}$ is also $(\theta, \rho, \epsilon, 2e^{\epsilon}\theta)$-robust.

This completes the proof of Theorem 2. □

C Omitted Details from Section 5

We will now define a new mechanism to get a new histogram from a given histogram in a differential private manner.

Mechanism (Restating Mechanism 4)**.** *For each bar of the histogram, independently add a noise according to the following probability distribution:*

$$\pi_q(x) = \begin{cases} \frac{1}{1-e^{-\frac{\sqrt{q}}{2}}} Lap(x \mid -\frac{q}{2}, \sqrt{q}) & -q < x < 0 \\ 0 & elsewhere \end{cases}$$

and then round each of the bar to the nearest integer. If any of the bar becomes negative then set it to 0. Output the resulting histogram.

Lemma 5. *For any $q \geq 1$, if in Mechanism 4 we output the histogram before rounding and setting negative values to 0 then we get (ϵ, δ)-differential privacy where $\epsilon = \frac{1}{\sqrt{q}}$ and $\delta = \frac{4}{\sqrt{q}}e^{-\frac{\sqrt{q}}{2}}$.*

Proof. Let t be the size of domain of the values in the database. Let $\mathbf{x}$ and $\mathbf{x}'$ be two neighbouring databases and let n_k and n'_k represent the number of values of k^{th} type in $\mathbf{x}$ and $\mathbf{x}'$ respectively. Let i^* be the index of the type in which $\mathbf{x}$ and $\mathbf{x}'$ differ. That means that $n_k = n'_k$ if $k \neq i^*$ and $n_{i^*} = n'_{i^*} + 1$. $p_D(s)$ denotes the probability of outputting s from a distribution D.

$$\Pr[\mathcal{M}(\mathbf{x}) \in S] = \int_{s\in S} p_{\mathcal{M}(\mathbf{x})}(s)\cdot ds$$
$$= \int_{s\in S} \big[\prod_{i=1}^{t} \pi_q(s_i - n_i)\big]\cdot ds \tag{21}$$
$$\Pr[\mathcal{M}(\mathbf{x}') \in S] = \int_{s\in S} \big[\prod_{i=1}^{t} \pi_q(s_i - n'_i)\big]\cdot ds \tag{22}$$

Using the fact that $\forall k \neq i^*, n_k = n'_k$ and $n_{i^*} = n'_{i^*} + 1$, we can divide S into 6 sets as follows:
S_0 contains those outputs in which $-\frac{q}{2} \leq s_{i^*} - n_{i^*}, s_{i^*} - n'_{i^*} < 0$,
S_1 contains those outputs in which $-q < s_{i^*} - n_{i^*}, s_{i^*} - n'_{i^*} \leq -\frac{q}{2}$,
S_2 contains those outputs in which $-q < s_{i^*} - n_{i^*} < -\frac{q}{2} < s_{i^*} - n'_{i^*} < 0$,
S_3 contains those outputs in which $s_{i^*} - n_{i^*} \leq -q < s_{i^*} - n'_{i^*} < 0$,
S_4 contains those outputs in which $-q < s_{i^*} - n_{i^*} < 0 \leq s_{i^*} - n'_{i^*}$ and
S_5 contains those outputs in which either $0 \leq s_{i^*} - n_{i^*}, s_{i^*} - n'_{i^*}$ or $s_{i^*} - n_{i^*}, s_{i^*} - n'_{i^*} \leq -q$.

For each of the above set except S_4, we analyze the difference between $a(= \pi_q(s_{i^*} - n'_{i^*}))$ and $b(= \pi_q(s_{i^*} - n_{i^*}) = \pi_q(s_{i^*} - n'_{i^*} - 1))$.
For S_0, $b = ae^{\frac{1}{\sqrt{q}}} \leq ae^{\epsilon}$
For S_1, $b = ae^{-\frac{1}{\sqrt{q}}} \leq ae^{\epsilon}$
For S_2, $b \leq ae^{\frac{1}{\sqrt{q}}} \leq ae^{\epsilon}$
For S_3, $b = 0 \leq ae^{\epsilon}$
For S_5, $b = 0 \leq ae^{\epsilon}$

Now let us analyze (21) for each S_i:

Case1: $S_i, i = 0, 1, 2, 3, 5$

$$\begin{aligned}
&= \int_{s\in S_i} \big[\prod_{i=1}^{t} \pi_q(s_i - n_i)\big]\cdot ds \\
&= \int_{s\in S_i} \big[\prod_{\substack{i=1\\ i\neq i^*}}^{t} \pi_q(s_i - n_i)\big]\pi_q(s_{i^*} - n_{i^*})\cdot ds \\
&= \int_{s\in S_i} \big[\prod_{\substack{i=1\\ i\neq i^*}}^{t} \pi_q(s_i - n'_i)\big]\pi_q(s_{i^*} - n'_{i^*} - 1)\cdot ds \\
&\leq \int_{s\in S_i} \big[\prod_{\substack{i=1\\ i\neq i^*}}^{t} \pi_q(s_i - n'_i)\big]\pi_q(s_{i^*} - n'_{i^*})e^{\epsilon}\cdot ds \\
&= e^{\epsilon}\int_{s\in S_i} \big[\prod_{i=1}^{t} \pi_q(s_i - n'_i)\big]\cdot ds
\end{aligned}$$

Case2: S_4, the condition forces $s_{i^*} - n_{i^*}$ to be in $[-1, 0)$

$$\begin{aligned}
&= \int_{s\in S_4} \big[\prod_{i=1}^{t} \pi_q(s_i - n_i)\big]\cdot ds \\
&= \int_{s_1} \dots \int_{s_{i^*}} \dots \int_{s_t} \big[\prod_{i=1}^{t} \pi_q(s_i - n_i)\big]\cdot ds_t \dots ds_{i^*} \dots ds_1
\end{aligned}$$

$$
\begin{aligned}
&= \int_{s_{i^*}} \pi_q(s_{i^*} - n_{i^*}) \Big(\int_{s_1} \dots \int_{s_t} [\prod_{\substack{i=1 \\ i \neq i^*}}^{t} \pi_q(s_i - n_i)] \cdot ds_t \dots ds_1 \Big) ds_{i^*} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(as } s_{i^*} \text{ varies from -1 to 0 independent of other } s'_k s) \\
&\leq \int_{s_{i^*}} \pi_q(s_{i^*} - n_{i^*})(1) ds_{i^*} \\
&= \int_{-1}^{0} \pi_q(s_{i^*} - n_{i^*}) ds_{i^*} \\
&= \frac{e^{1/\sqrt{q}} - 1}{2(1 - e^{-\sqrt{q}/2})} e^{-\sqrt{q}/2} \\
&\leq 2(e^{1/\sqrt{q}} - 1) e^{-\sqrt{q}/2} && (1 - e^{-\sqrt{q}/2} \geq 1 - e^{-1/2} \geq \tfrac{1}{4}) \\
&\leq \frac{4}{\sqrt{q}} e^{-\sqrt{q}/2} && (\text{For } x \leq 1, e^x \leq 2x + 1) \\
&= \delta
\end{aligned}
$$

Now we combine the above results with Equations (21) and (22) as follows:

$$
\begin{aligned}
\Pr[\mathcal{M}(\mathbf{x}) \in S] &= \int_{s \in S} p_{\mathcal{M}(\mathbf{x})}(s) \cdot ds \\
&= \sum_{l=0}^{5} \int_{s \in S_l} p_{\mathcal{M}(\mathbf{x})}(s) \cdot ds \\
&= \sum_{\substack{l=0 \\ l \neq 4}}^{5} \int_{s \in S_l} p_{\mathcal{M}(\mathbf{x})}(s) \cdot ds + \int_{s \in S_4} p_{\mathcal{M}(\mathbf{x})}(s) \cdot ds \\
&\leq \sum_{\substack{l=0 \\ l \neq 4}}^{5} e^{\epsilon} \int_{s \in S_l} p_{\mathcal{M}(\mathbf{x}')}(s) \cdot ds + \delta \\
&= e^{\epsilon} \Big(\sum_{\substack{l=0 \\ l \neq 4}}^{5} \int_{s \in S_l} p_{\mathcal{M}(\mathbf{x}')}(s) \cdot ds \Big) + \delta \\
&= e^{\epsilon} \Big(\sum_{\substack{l=0 \\ l \neq 4}}^{5} \int_{s \in S_l} p_{\mathcal{M}(\mathbf{x}')}(s) \cdot ds + 0 \Big) + \delta \\
&= e^{\epsilon} \Big(\sum_{\substack{l=0 \\ l \neq 4}}^{5} \int_{s \in S_l} p_{\mathcal{M}(\mathbf{x}')}(s) \cdot ds + \int_{s \in S_4} p_{\mathcal{M}(\mathbf{x}')}(s) \cdot ds \Big) + \delta \\
&= e^{\epsilon} \Big(\sum_{l=0}^{5} \int_{s \in S_l} p_{\mathcal{M}(\mathbf{x}')}(s) \cdot ds \Big) + \delta \\
&= e^{\epsilon} \Pr[\mathcal{M}(\mathbf{x}') \in S] + \delta
\end{aligned}
$$

Similarly we can also prove the following: $\Pr[\mathcal{M}(\mathbf{x}') \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(\mathbf{x}) \in S] + \delta$. □

Theorem 8. *For any $q \geq 1$, Mechanism 4 achieves (ϵ, δ)-differential privacy where $\epsilon = \frac{1}{\sqrt{q}}$ and $\delta = \frac{4}{\sqrt{q}} e^{-\frac{\sqrt{q}}{2}}$.*

Proof. Rounding and setting negative values to 0 is a post-processing step after adding noise.By Lemma 5, the addition of noise is (ϵ, δ)-private . Using the fact from [1] that post-processing doesn't decrease the privacy, we can say that Mechanism 4 is also (ϵ, δ) private. □

Theorem (Restating Theorem 4). *Mechanism 4 achieves $(\alpha, 0, 0)$ accuracy with $\alpha = qt$.*

Proof. This can be easily seen. In essence the histogram mechanism deletes some elements from the buckets. Now, the mechanism deletes almost q elements from a bucket, as the probability of adding more negative noise is zero. Hence, the histogram returned by the mechanism can be obtained by deleting at most qt, where t is the number of buckets from the original database. Thus, by the considering the distorted random variable as the corresponding database for every instance of the histogram mechanism noise, we get its distribution as that of output distributions of the mechanism. Hence, the mechanism is $(qt, 0, 0)$ accurate. □

D Omitted Details from Section 6

Now, we present our privacy composition theorem.

Theorem (Restating Theorem 6). *Let $M_1 : A \to B$ be a neighbourhood preserving mechanism for a function $f_1 : A \to B$, and let $M_2 : B \to C$ be a mechanism which is (ϵ, δ)-differential private for a function $f_2 : B \to C$. Then the composite mechanism, $M_2 \circ M_1 : A \to C$ is (ϵ, δ)-differential private.*

Proof. Note that we defined a randomized mechanism to be neighbourhood preserving, if it is a convex combination of neighbourhood preserving deterministic functions.

First we consider the case when the mechanism M_1 is deterministic. In this case, for neighbouring databases $\mathbf{x}, \mathbf{x}'$, $M_1(\mathbf{x}), M_1(\mathbf{x}')$ are deterministic databases in B, which are also neighbours. Hence, by the (ϵ, δ)-differential privacy of the mechanism M_2, for all subsets $S \subseteq C$, we have

$$P(M_2(M_1(\mathbf{x})) \in S) \leq e^{\epsilon} P(M_2(M_1(\mathbf{x}')) \in S) + \delta.$$

This proves the privacy guarantee of the mechanism in case of deterministic M_1.

For randomized mechanisms M_1, we use the fact that we can represent a randomized mechanism as a convex combination of deterministic mechanisms, where the linear weights corresponds to the randomness in the mechanism. Thus, we have $M_1 = \sum_{i=1}^{n} p_i M_{1,i}$, where $\sum_{i=1}^{n} p_i = 1$. Here, $M_{1,i}$ are the deterministic mechanism, and $M_1 = M_{1,i}$ with probability p_i. This implies that for every $S \subseteq C$, we have

$$P(M_1(\mathbf{x}) \in S) = \sum_{i=1}^{n} p_i P(M_{1,i}(\mathbf{x}) \in S).$$

Now, for two neighbouring databases $\mathbf{x}, \mathbf{x}' \in A$, and any subset $S \subseteq C$, we have

$$\begin{aligned} P(M_2(M_1(\mathbf{x})) \in S) &= \sum_{i=1}^{n} p_i P(M_2(M_{1,i}(\mathbf{x})) \in S) \\ &\overset{\text{(a)}}{\leq} \sum_{i=1}^{n} p_i (e^{\epsilon} P(M_2(M_{1,i}(\mathbf{x}') \in S) + \delta) \\ &= e^{\epsilon} \sum_{i=1}^{n} p_i P(M_2(M_{1,i}(\mathbf{x}') \in S) + \delta \\ &= e^{\epsilon} P(M_2(M_1(\mathbf{x}') \in S) + \delta \end{aligned}$$

Here (a) follows because M_2 is (ϵ, δ) private and $M_{1,i}(\mathbf{x})$ and $M_{1,i}(\mathbf{x}')$ are neighbours in B. This proves that the (ϵ, δ) privacy guarantee of the composite mechanism $M_2 \circ M_1$. □

Now, we show how appropriate mechanisms can be composed to achieve good accuracy and privacy guarantees for computing a statistic on a database. We consider databases which consists of positive real number numbers. The bucketing is done with t number of buckets, i.e, rounding each element of database to the nearest multiple of $\frac{B}{t}$.

Lemma (Restating Lemma 4). *The bucketing followed by Mechanism 4 is an $(\alpha_p, \beta_p, 0)$-accurate and (ϵ_p, δ_p)-differentially private mechanism for the identity function, with $\alpha_p = qt, \beta_p = \frac{B}{2t}, \epsilon_p = \frac{1}{\sqrt{q}}, \delta_p = \frac{4}{\sqrt{q}} e^{-\frac{\sqrt{q}}{2}}$.*

Proof. Let us denote the Bucketing mechanism by $\mathcal{M}_{buck}$ and Mechanism 4 by $\mathcal{M}_{hist}$. We compute distance between histograms by considering them as distributions and taking the Wasserstein distance between them. Now, using this distance metric, we have the bucketing mechanism to be $(0, \frac{B}{2t}, 0)$ for identity function, as each element is distorted by at most $B/2t$, which means that the corresponding histogram will shift by at most the same distance.

Also note that bucketing mechanism is a deterministic mechanism that is neighbourhood preserving. This is because, removing a single element changes the output of bucketing by at most one element. Hence, neighbours remain neighbours after bucketing. This implies that the distortion sensitivity function of bucketing mechanism is identity, i.e, $\sigma_{\mathcal{M}_{buck}}(\alpha') = \alpha'$.

We also have that the histogram mechanism is $(qt, 0, 0)$-accurate (see Theorem 4) and $(\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}}e^{-\frac{\sqrt{q}}{2}})$-differentially private (see Theorem 8). Note that the error sensitivity function of the histogram mechanism is also the identity function, i.e, $\tau_{\mathcal{M}_{hist}}(\beta') = \beta'$. This follows from Claim 1, which states that adding the same noise doesn't change the wasserstein distance between distributions.

Hence, by application of Theorem 5 and the fact that post-processing preserves differential privacy [1, Proposition 2.1], we have that the bucketing followed by histogram mechanism is $(qt, \frac{B}{2t}, 0)$-accurate and $(\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}}e^{-\frac{\sqrt{q}}{2}})$-differentially private.

□

D.1 Case Study

We now show the application of the histogram for computing Max and support of a database in an accurate and private manner.

D.1.1 Maximum

Let d be a distance metric on $\mathcal{X}$ which has a well-defined order. We define the function $f : \mathcal{N}^{\mathcal{X}} -> \mathcal{X}$, as one which takes a takes a database as input and outputs the largest element and take d_1 as the maximum d-distance each element of a histogram is to be modified to get the another, d_2 as d-distance in $\mathcal{X}$ space. Note that f is a perfectly accurate mechanism for maximum. Also the error-sensitivity function for f is identity, i.e, $\tau_f(\beta') = \beta'$. Because if the input distributions of histograms have a wasserstein d_1-wasserstein β, implies that the maxima of these distributions are also distributions which are atmost d_2-wasserstein β apart.
Now, using 5 and 6, we have that, the histogram mechanism for finding max is $(\alpha_p, \beta_p, 0)$ accurate and (ϵ_p, δ_p) private.

D.1.2 Support

Here we define the function $f : \mathcal{N}^{\mathcal{X}} -> \mathcal{N}^{\mathcal{X}}$, which takes a takes database as input and outputs the database after removing duplicates and take d_1, d_2 as wasserstein distance between inputs interpreting the histograms as distributions. Again, this is a perfectly accurate and private mechanism for maximum. Also the error-sensitivity function for f is identity, i.e, $\tau_f(\beta') = \beta'$. This is because, by the same above analysis and if each element is modified by atmost β', then the support elements also changes by atmost β'.
Now, using 5 and 6, we have that, the histogram mechanism for finding max is $(\alpha_p, \beta_p, 0)$ accurate and (ϵ_p, δ_p) private.