Bias Reduction in Differentially Private Quantile Estimation

Aymane Moataz, Jan Ramon *

Abstract

Quantile release mechanisms are vital for differentially private data analysis, yet their utility is often constrained by the existing bias in current implementations based on the exponential mechanism. In this work, we propose strategies to mitigate this bias, achieving significant improvements in the balance between privacy and utility for both single and multiple quantile releases. Our approach includes more efficient methods for dividing the privacy budget to reduce conditional bias, demonstrating that fully eliminating this bias requires a quadratic expansion of the clipping range. Furthermore, our adaptation of the exponential mechanism enables its application to a case study, where the inherent randomness in Gaussian data can be used to guarantee differential privacy in quantile release.

1 Introduction

The rapid expansion of data-driven applications across domains such as healthcare, finance, and social sciences has raised pressing privacy concerns surrounding the analysis of sensitive information. Differential Privacy [DMNS06] has emerged as a foundational framework for addressing these concerns, offering robust guarantees by ensuring that the inclusion or exclusion of any individual data point has minimal impact on analysis results. This framework has gained widespread adoption in industries [EPK14, App17, DKY17], research institutions, and government agencies [MSI23], empowering organizations to extract insights from sensitive data while preserving individual privacy. Applications of DP range from private user behavior analysis in technology companies to the protection of sensitive medical and census data.

Summary statistics, such as means, variances, and percentiles, play a crucial role in synthesizing large datasets for decision-making across sectors. However, directly computing these statistics on sensitive data can lead to privacy breaches, as even aggregate values may reveal information about individuals.

In this paper, we develop methods that improve the utility of differentially private single quantile release. Efficient many-quantile algorithms often depend

*MAGNET Team, INRIA Lille, France

on robust mechanisms for single quantile release [KSS22], making advancements in this area broadly impactful.

Existing methods for releasing differentially private quantiles face two main challenges: biases in implementations of the exponential mechanism and inefficient allocation of the privacy budget. These limitations hinder their ability to achieve optimal trade-offs between privacy and utility in real-world scenarios.

The exponential mechanism, defined in Definition 2.5, operates within a specified clipping range $[a, b] \subset \mathbb{R}$ and uses a utility function u to assign scores to values in the interval $[a, b]$ utilizing the private dataset. This mechanism ensures that values with lower scores are exponentially less likely to be selected. It is particularly useful for releasing values under differential privacy (DP) constraints, especially when the scores assigned by u exhibit only minor changes in response to slight modifications in the dataset.

In such cases, the exponential mechanism generally outperforms alternatives like the Laplace F.1 or Gaussian mechanisms F.2 by minimizing noise while preserving utility.

A common method for releasing quantiles [Smi11, AMS+20, DGHM+22, KSS22], involves using a utility function that remains constant over a set of intervals $\mathcal{I} = \{I_k\}_{k\in[n-1]}$. The private quantile estimator $\mathcal{M}_q(X)$ is sampled according to the following probability density function:

$$f_{\mathcal{M}_q(X)}(o) = \begin{cases} \sum_{I_k\in\mathcal{I}} \frac{1}{|I_k|} \mathbb{1}_{I_k}(o) \frac{|I_k| u(I_k)}{\int_a^b u(s,x)\, ds} & \text{if } o \in [a, b], \\ 0 & \text{otherwise.} \end{cases}$$

Typically, the intervals $\mathcal{I}$ are defined based on the ordered samples $x_{(.)}$, with $I_k = (x_{(k)}, x_{(k+1)}]$. However, this method introduces bias.

$$\mathbb{E}(\mathcal{M}_q(X)) \neq x_{(\lceil qn \rceil)}.$$

The probability of selecting points from an interval I_m is proportional to its length and score:

$$\int_{s\in I_m} f_{\exp}(s)\, ds \propto (x_{(m+1)} - x_{(m)}) u(I_m).$$

As a result, larger intervals are more likely to be selected, which introduces additional bias.

Even under ideal conditions—when the selected interval $I_k = (x_{(k)}, x_{(k+1)}]$ matches the intended quantile—the mechanism outputs a value uniformly distributed within $[x_{(\lceil qn \rceil)}, x_{(\lceil qn \rceil+1)})$. Despite this, the mechanism remains conditionally biased, as its expected value is skewed even when conditioned on selecting the highest utility interval, denoted by the event E_m:

$$\mathbb{E}(\mathcal{M}_q(X) \mid E_m) \neq x_{(\lceil qn \rceil)}.$$

To address these biases, we propose methods to reduce both the unconditional and conditional bias in single quantile release mechanisms. Additionally, we aim

to improve the efficiency of privacy budget allocation to enhance the utility of both single and multiple quantile release mechanisms. Our contributions are

- Designing conditionally unbiased quantile release mechanisms using various sampling strategies (Uniform, Gaussian, Laplace), with new hyperparameters that distribute more probability mass near the target quantile [1] 4.
- Applying the new sampling techniques to estimate quantiles when data is perturbed wth Gaussian noise 5.
- Providing theoretical motivations for the approaches, to guide hyperparameter selection 6.
- Demonstrating that fully eradicating bias in DP quantile release is constrained by a 'no free lunch' theorem, resulting in a quadratic expansion of the clipping range7.
- Empirically evaluating the mechanisms and comparing their performance with prior work 8.

The structure of the paper is as follows: Section 3 reviews related work and highlights key limitations in existing methods. Sections 4, 5 detail our proposed methods for reducing conditional bias and using Gaussian data, respectively. Section 6 establishes the theoretical guarantees of the proposed methods, section 7 introduces a fully unbiased mechanism. Section 8 presents experimental results, showcasing the performance of our methods compared to state-of-the-art approaches. Finally, Section 9 explores the implications of our findings, outlines limitations, and suggests directions for future research.

To promote reproducibility and facilitate practical adoption, we provide an open-source implementation of our methods, available at `DP_Quantiles`. The repository contains all the necessary code to replicate our results and apply the proposed techniques to new datasets.

2 Preliminaries

In the following, we fix $n \in \mathbb{N}$ number of samples in the private dataset and $q \in (0, 1)$ the target quantile, $r_q = \lceil qn \rceil + 1$. The notation $[n]$ will be used to represent the set of the first n strictly positive integers. We define the following notations:

- y_i denotes the i-th entry of the vector $Y \in \mathbb{R}^n$,
- $y_{(i)}$ denotes the i-th smallest entry in Y, i.e., the i-th order statistic, where the entries of Y are sorted in ascending order,

[1] Conditioned on selecting the highest utility interval, the proposed estimators are unbiased for $x_{(\lceil qn \rceil)}$.

- $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2.
- $\mathcal{TN}(\mu, a, b, \sigma^2)$ denotes a truncated normal distribution in $[a, b]$ with mean μ and variance σ^2.
- Let $a, b \in \mathbb{R}$. We use $|\cdot|$ to denote the length of an interval. In particular, $|[a, b]| = b - a$.

We now proceed to define key concepts such as the sample quantile, adjacency between datasets and differential privacy.

Definition 2.1 (Sample Quantile)**.** Let $X \in \mathbb{R}^n$, drawn i.i.d from an unknown distribution P. We define the sample quantile is defined as

$$Q_q(X) = x_{(\lceil qn \rceil)}.$$

If the distribution P has a density function f, and a cumulative distribution function (CDF) F.

Using Slutsky's Lemma and the Central Limit Theorem, it can be shown that

$$Q_q(X) \xrightarrow{d} \mathcal{N}\left(F^{-1}(q), \frac{1}{4nf\left(F^{-1}(q)\right)^2}\right).$$

where $\xrightarrow{d}$ denotes the convergence in distribution as the smaple size n gets large.

Definition 2.2 (Adjacency of Datasets)**.** Two vectors $X, X' \in \mathbb{R}^n$ are said to be adjacent under the "replacing neighboring" relation if and only if there exists exactly one index $j \in [n]$ such that:

$$X_j \neq X'_j.$$

In the following, we will use the notation $X \sim X'$ to denote that the datasets X and X' are adjacent according to the definition of neighboring datasets under replacement.

Definition 2.3 (Sensitivity of a Function)**.** Let $f : \mathcal{D} \to \mathbb{R}^d$ be a function defined on a domain $\mathcal{D} \subseteq \mathbb{R}^n$. The sensitivity Δf of f is the maximum change in the value of f when a single data point in the dataset is modified. More formally, the sensitivity is defined as:

For $p \in \{1, 2\}$, we write:

$$\Delta_p f = \max_{D \sim D' \in \mathcal{D}} ||f(D) - f(D')|\,|_p,$$

where D and D' are adjacent datasets (differing by exactly one data point), and $f(D)$ denotes the value of the function applied to dataset D.

We also denote $\Delta = \Delta_1$ for convenience.

Definition 2.4 (Differential Privacy [DMNS06])**.** Let $M : \mathbb{R}^n \to \mathbb{R}^d$ be a mechanism. We say that M satisfies (ϵ, δ)-differential privacy if, for any two adjacent datasets $D \sim D' \in \mathcal{D} \subseteq \mathbb{R}^n$ and any Borel set $O \in \mathcal{B}(\mathbb{R}^d)$, the following condition holds:

$$\mathbb{P}(M(D) \in O) \leq e^{\epsilon}\mathbb{P}(M(D') \in O) + \delta.$$

This property is known as (ϵ, δ)-indistinguishability. If $\delta = 0$, the mechanism M is said to satisfy pure differential privacy.

Next, we introduce the exponential mechanism, a well-known mechanism that satisfies pure differential privacy.

Definition 2.5 (Exponential Mechanism [MT07])**.** Given a dataset $D \in \mathcal{D} \subseteq \mathbb{R}^n$, a finite range $\mathcal{R} \subseteq \mathbb{R}^d$, a utility function $u : \mathcal{R} \times \mathcal{D} \to \mathbb{R}$ that assigns scores to each outcome $r \in \mathcal{R}$ based on the dataset D, and a privacy parameter ϵ, the exponential mechanism selects and outputs an element $r \in \mathcal{R}$ with probability proportional to:

$$\exp\left(\frac{\epsilon u(r, D)}{2\Delta u}\right),$$

where Δu is the sensitivity of the utility function 2.3

3 Related work

Differentially private single quantile release:

The problem of privately computing and releasing quantiles has received considerable attention in the literature [NRS07, DL09, Smi11, AD20, DGHM+22, KSS22, LGG23a]. A common approach involves adding noise proportional to the maximum possible variation in the dataset, which can be substantial when the data contains extreme values or covers a wide range.

To address this, Nissim et al. [NRS07] proposed a technique called "smooth sensitivity," which adapts to the data itself by adjusting noise based on the particular characteristics of the dataset. Similarly, Dwork et al. [DL09] developed the "propose-test-release" approach, where the algorithm proposes a database-dependent noise scale and verifies its feasibility. In this case, an estimate is released; if not, the algorithm withholds the result to avoid risking privacy. Although effective, this approach does not always provide a value, both of these methods guarantee approximate differential privacy.

More recent methods focus on a probabilistic framework known as the exponential mechanism [MT07]. This method allows for more precise quantile estimation by setting up a probability distribution over possible outcomes that is shaped by how closely the values are to the target quantile. For example, Asi et al. [AD20] discuss a technique for estimating the median based on this principle, and Drechsler et al. [DGHM+22] used this approach to construct confidence intervals for the median.

Differentially private multiple quantile release:

A straightforward approach to releasing multiple quantiles is to estimate each one individually and then combine the results. However, this can reduce accuracy, as the privacy cost grows polynomially with the number of releases.

An alternative approach involves releasing a private cumulative distribution function (CDF), which can provide multiple quantiles at once. However, this can often disclose more information than is necessary for many applications, especially when the number of quantiles is small, as demonstrated in [LGG23a]. This can lead to the need for a higher privacy budget.

To mitigate these issues, recent methods [GJK21, KSS22, LGG23a] have been introduced to better balance privacy and accuracy. The current leading approach [KSS22] with comparable performance to [LGG23a], repeatedly applies single quantile mechanisms using tree aggregation techniques [CSS11, DRV10]. This approach suggests that improvements in the release of a single quantile mechanisms can result in significant benefits when releasing multiple quantiles.

Data perturbation and differential privacy:

An intuitive approach to protecting individual privacy involves modifying database entries, either by removing identifying information [Swe02] or by adding random noise. In the realm of differential privacy, a similar concept called local differential privacy (LDP) guarantees privacy for individual data records by ensuring that any operation on these records preserves differential privacy. However, since LDP is not specifically designed for particular mechanisms, it may offer more privacy protection than needed in many use cases.

When noisy measurements are already available, it becomes valuable to explore how these can be integrated with differential privacy techniques. This paper investigates the connection between randomized data and differential privacy, aiming to leverage randomized data for better privacy accounting and improved utility.

Building on the non-uniform sampling strategies developed for single quantile release, we can optimize the privacy-utility tradeoff in scenarios involving Gaussian data.

Previous work, such as [JLY+24] also explored the connection between randomized data and differential privacy, in the case of linear regression, however they used the perturbed data to guarantee local differential privacy.

4 Centered differentially private quantile release mechanisms

In this section, we introduce a new mechanism designed to allocate probability mass differently than the traditional exponential mechanism used for differentially private quantile release. This mechanism introduces a novel set of constant-sized intervals centered around the target samples, aiming to reduce the conditional bias present in current implementations. Additionally, it minimizes the influence of large intervals by incorporating more intervals around the data points, while

also introducing additional parameters that allocate more mass to the highest utility interval.

4.1 Centered Exponential Mechanism - CEXP

The first mechanism, which we refer to as the centered exponential mechanism, employs a uniform distribution to sample from chosen intervals and also involves preprocessing the dataset. A formal definition of this mechanism is provided below.

Definition 4.1 (Centered Exponential Mechanism)**.** Let $q \in (0,1)$, $a, b \in \mathbb{R}$, $\omega > 0$, and $X^O \in [a,b]^n$ be a dataset.

We construct an extended dataset $X^o \in [a,b]^{n+2}$ by adding boundary values:

$$X^o_{(1)} = a, \quad X^o_{(n+2)} = b.$$

This changes the rank of the sample quantile we want to estimate to $r_q = \lceil qn \rceil + 1$ Next, we create a new vector X by ensuring a distance of at least $2\omega > 0$ between samples.

Formally, for $k \geq 1$, we define:

$$\begin{aligned} X_{(r_q+k)} &= X_{(r_q+k-1)} + t^+(k), \\ X_{(r_q-k)} &= X_{(r_q-k+1)} + t^-(k). \end{aligned}$$

where $t^+(k) = \max\left(2\omega, X^o_{(r_q+k)} - X_{(r_q+k-1)}\right)$ and $t^-(k) = -\max\left(2\omega, X_{(r_q-k+1)} - X^o_{(r_q-k)}\right)$

The dataset creates a range $\mathcal{R}_X :\subseteq \mathcal{R} := [a-(\lceil qn \rceil - 1)\omega, b+(n-\lceil qn \rceil - 1)\omega]$ into $2n+3$ intervals, denoted as I_k, defined as follows:

$$I_k = \begin{cases} [x_{(k)} - \omega, x_{(k)} + \omega], & \text{if } k \text{ is odd,} \\ [x_{(k)} + \omega, x_{(k)} - \omega], & \text{if } k \text{ is even.} \end{cases}$$

Let E_m denote the event of selecting an interval I_m from the range $\mathcal{R}_X$.

The mechanism generates an output $o \in \mathcal{R}_X$ with the following probability density:

$$f(o) = \sum_{I_k \in \mathcal{R}_X} f_{I_k}(o) \cdot \mathbb{P}(E_k),$$

where the density function $f_{I_k}(o)$ is determined by the interval I_k:

$$f_{I_k}(o) = \begin{cases} \frac{1}{2\omega} \mathbb{1}_{I_k}(o), & \text{if } k \text{ is odd,} \\ \frac{1}{x_{(k+1)} - x_{(k)} - 2\omega} \mathbb{1}_{I_k}(o), & \text{if } k \text{ is even.} \end{cases}$$

The probability of selecting interval E_k is given by:

$$\mathbb{P}(E_k) \propto \begin{cases} (x_{(k+1)} - x_{(k)} - 2\omega) \cdot u(I_k), & \text{if } k \text{ is even,} \\ 2\omega \cdot u(I_k), & \text{if } k \text{ is odd.} \end{cases}$$

The utility function $u(I_k)$ is defined as:

$$u(I_k) = \begin{cases} \exp(\epsilon_\mu) \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is odd and } r(k) = 0, \\ \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is odd and } r(k) \neq 0, \\ \exp\left(-\frac{\epsilon_a}{2}\right) \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is even.} \end{cases}$$

Here, $r(k)$ is the rank function, defined as:

$$r(k) = \left| \left\lceil \frac{k}{2} \right\rceil - r_q \right|,$$

where:

- $u(I_k)$ is the utility function for interval I_k,
- $r(k)$ is the rank of the sample relative to the target quantile.

The parameter ϵ_a is chosen to reduce the probability mass assigned to the larger intervals $[x_{(i)} + 2\omega, x_{(i+1)} - 2\omega]$, and ϵ_μ is used to enhance the probability mass of the interval centered around the target quantile $[x_{(r_q)} - \omega, x_{(r_q)} + \omega]$. Essentially, the goal is to adjust the distribution of the exponential mechanism so that actual data points are more likely to be sampled. This can be understood as dedicating part of the overall privacy budget to minimize the probability of selecting values from these less relevant intervals.

Remark 4.2. in the remainder of the paper , We will consider the adjacent dataset is created by changing a value to a value above it (Similar results can be obtained in other cases due the symmetry of the proposed method, it suffices to exchange X, X'). We define the quantity $\text{Gap}_q = x'_{(\lceil qn \rceil)} - x_{(\lceil qn \rceil)} \leq x_{(\lceil qn \rceil + 1)} - x_{(\lceil qn \rceil)}$

Lemma 4.3. *Let $X \in \mathcal{R}_X$ constructed following 4 and X' constructed similarly from adjacent dataset $X^{O,'}$ by changing a value to a larger value we have,*

for all $k \in [0, n - r_q + 1]$

$$x_{(r_q+k)} \leq x'_{(r_q+k+1)}$$
$$x'_{(r_q+k)} \leq x_{(r_q+k+1)}$$

for all $k \in [0, r_q - 2]$

$$x_{(r_q-k-1)} \leq x'_{(r_q-k)}$$
$$x'_{(r_q-k-1)} \leq x_{(r_q-k)}$$

Theorem 4.4 (Centered exponential mechanism DP guarantees)**.** *Sampling an element in the interval $\mathcal{R}_X$ with the density defined in 4.1 satisfies pure $\epsilon_\mu + \epsilon_a + \epsilon$ differential privacy.*

The proof, presented in Appendix A.2.1, The idea of the proof is to create a subdivision $\{d\}$ of the interval $[a, b]$ that isolates the union $U := I_{2\lceil qn \rceil - 1} \cup I'_{2\lceil qn \rceil - 1} := [d_j, d_{j+1}] \cup [d_{j+1}, d_{j+2}] \cup [d_{j+2}, d_{j+3}]$ for some index j under which the utility function, where the central interval $[d_{j+1}, d_{j+2}] = I'_{2\lceil qn \rceil - 1}$ to further isolate the highest utility interval. Using this technique we can bound the normalization ratio without using ϵ_μ and thus avoid dividing the budget ϵ_μ by two.

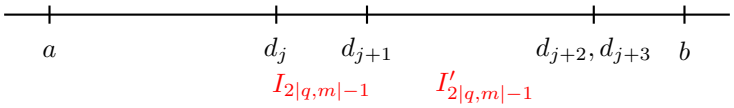

Figure 1: Illustration of intervals and points used in the proof

Example 4.5. We will give an example where the centered exponential mechanism significantly outperforms the classical exponential mechanism, we will select a large gap between the target quantile and the next datapoint, this also causes a high probability density far from the target for the centered exponential mechanism but it is attenuated by the factor $\exp -\epsilon_a$, for a total budget $\epsilon = 5$ we allocate an attenuation budget $\epsilon_a = 2$

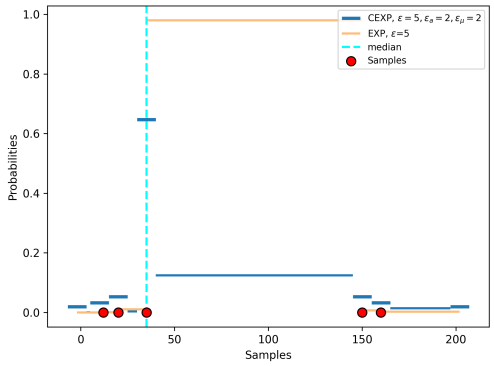

Figure 2: Comparaison between The two distributions, $n = 5 \quad \epsilon = 5 \qquad [a, b] = [0, 200] \quad \omega = 5$.

4.2 Gaussian Exponential mechanism - GEXP

Definition 4.6 (Gaussian Exponential mechanism)**.** Let X, ω, n, $\{E_k\}_{k \in [2n+3]}$, and $\{I_k\}_{k \in [2n+3]}$ be defined as in section 4. Compared to the centered exponential mechanism, we sample from the constant-size intervals using a truncated normal

distribution:

$$f(o \mid E_{2k-1}) = f_{\mathcal{TN}\left(x_k, x_k-\omega, x_k+\omega, \sigma^2_{GEXP}\right)} \mathbb{1}_{I_{2k-1}}(o)$$

A full version of this definition is provided in the appendix A.1

Theorem 4.7 (Gaussian Exponential Mechanism DP Guarantees)**.** *Let* $\sigma_{GEXP} > 0$*. Sampling an element in the range* $\mathcal{R}_X \subseteq [a-(\lceil qn \rceil - 1)\omega, b+(n-\lceil qn \rceil - 1)\omega]$ *with the density defined in Definition 4.6 satisfies pure differential privacy with*

$$\max \begin{pmatrix} \epsilon_a + \epsilon_\mu + \epsilon + \log(\dfrac{2\omega}{W}) \\ \dfrac{\omega^2}{2\sigma^2_{GEXP}} + \epsilon_\mu + \dfrac{\epsilon_a}{2} + \epsilon \\ \dfrac{\omega^2}{2\sigma^2_{GEXP}} + \epsilon + \log(\dfrac{W}{2\omega}) \\ \epsilon_a + \epsilon \end{pmatrix}$$

where $W = \sqrt{2\pi}\sigma_{GEXP}\left(\Phi\left(\frac{\omega}{\sigma_{GEXP}}\right) - \Phi\left(-\frac{\omega}{\sigma_{GEXP}}\right)\right)$.

Here, Φ *denotes the cumulative distribution function (CDF) of a standard normal distribution.*

It is also possible to use a truncated Laplace distribution, the definition of the privacy guarantees of the Laplace exponential mechanism are provided in the appendix. A.2

5 Gaussian Randomization

The gaussian exponential mechanism, can be very useful if the data is initially perturbed with Gaussian noise, this is a releastic case in many settings (time series data etc). We will assume in what follows that we can only access noisy data

We aim to develop a differentially private quantile estimator that is robust in the randomized data setting described below

Definition 5.1. Consider the set of sensitive values $D = \{x_i\}_{i=1}^n$ from a fixed but unknown distribution P. Consider the mechanism $\mathcal{M}_\sigma^{gauss}$ performing the following when applied to D:

- Measure each sensitive value, adding measurement noise with variance σ, i.e.,
$$\forall i \in [n] : \hat{x}_i = x_i + \eta_i \text{ with } \eta_i \sim \mathcal{N}(0, \sigma^2)$$

Our goal is to utilize the natural randomness in the observed samples to release quantiles while ensuring differential privacy. This inherent noise intuitively provides a better privacy-utility tradeoff compared to state-of-the-art differentially private quantile mechanisms. A key challenge is that these state-of-the-art

methods heavily rely on the exponential mechanism. However, as demonstrated in 4.2, we can adapt the exponential mechanism to work with Gaussian noise. In the following, we introduce further modifications to the Gaussian exponential mechanism to effectively leverage the noise present in the samples.

5.1 Randomized data simple differentially private quantile mechanism - RQM

Specifically, we propose to discretize the output space by creating intervals centered around the observed samples. Our approach involves locating the mean of the Gaussian distribution with high probability within these intervals. By sampling from a centered interval, we will demonstrate that releasing a sample from a normal distribution centered at any point within that interval preserves differential privacy.

This strategy allows us to guarantee approximate differential privacy while releasing data that has been perturbed with Gaussian noise, eliminating the need for additional randomization steps.

Definition 5.2 (Randomized data Quantile Mechanism DP Guarantees)**.** Consider the setting described in Section 4. Let $\mu_k \in I_{2k-1}$. Compared to the gaussian exponential mechanism we consider an arbitrary mean we use a normal distribution instead of a truncated normal:

$$f(o \mid E_k) = f_{\mathcal{N}(\mu_k,\sigma^2)} \mathbb{1}_{I_{2k-1}}(o)$$

A full version of this definition is provided in the appendix B.1

Theorem 5.3 (Randomized data Quantile Mechanism DP Guarantees)**.** *Let $\delta > 0$, choose ω s.t $\mathbb{P}_{\eta\sim\mathcal{N}(0,\sigma^2)}(|\eta| > \omega) \leq \delta$*

Sampling a value with the density defined in 5.2 satisfies approximate

$$\max \begin{pmatrix} \epsilon_a + \epsilon_\mu + \epsilon + \log(\frac{2\omega}{W}) \\ \frac{\omega^2}{2\sigma^2} + \epsilon_\mu + \frac{\epsilon_a}{2} + \epsilon \\ \frac{\omega^2}{2\sigma^2} + \epsilon + \log(\frac{W}{2\omega}) \\ \epsilon + \epsilon_a \end{pmatrix}, \delta$$

-differential privacy.

where $W = \sqrt{2\pi}\sigma$,

6 Consistency guarantees

We analyzed the reduction in conditional bias achieved by $CEXP$ and $GEXP$ and extended this investigation to study the convergence rates of our differentially

private quantile estimators. Our objective is to achieve reliable performance while addressing practical challenges arising from real-world data distributions.

To avoid imposing strong distributional assumptions, we introduce slight perturbations to the data by adding Gaussian noise. This approach aligns with the use case outlined in 5 and is supported by prior research [LGG+23b], which demonstrates the benefits of smoothing the output distribution when applying the exponential mechanism to peaked distributions. Common smoothing methods include convolving with a max kernel or adding continuous noise. By tackling the challenges associated with peaked distributions, we show that the consistency upper bounds, which often depend on a constant factor, can be substantially improved using the mechanisms introduced in Section 4.

Previous work on private medians, such as [AD20], has explored consistency guarantees for private quantile estimators but typically relies on two key assumptions: (1) that dP is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, and (2) that there is non-zero density in a region around the population quantile. However, in real-world scenarios, these assumptions may break down—for instance, even after perturbing the data, the density near the quantile may still be very small.

Our method circumvents these limitations by convolving P with a Gaussian distribution. This results in a smoothed distribution that is absolutely continuous with support on $\mathbb{R}$. Additionally, we allocate part of the privacy budget to reduce the constant factor in the consistency upper bound, thereby enhancing the robustness of our guarantees. Specifically, let q_s denote the sample quantile and $q_p \triangleq F^{-1}(q)$ the population quantile. For a real number $l_q > 0$, we define $d_q \triangleq \inf_{[q_s, q_s+t]} f(u)$, which represents the minimum density within $[q_s, q_s + t]$.

To derive consistency guarantees for the sample quantile, we apply the Hoeffding inequality and use the fact that

$$\mathbb{1}_{x \sim dP * G \leq q_p} \sim \text{Bernoulli}(n, F(q_p)).$$

Lemma 6.1. *For any $q \in (0,1)$ The following consistency result holds for the sample quantile as defined in 2.1*

$$\forall t > 0 \quad \mathbb{P}(|q_s - q_p| > t) \leq 2e^{-2nt^2 d_q^2}$$

Lemma 6.2. *Let $\omega > 0$ be defined as in 4. Fix $t > \omega$, and define $d_t \triangleq \inf_{u \in [q_s, q_s+t]} f(u)$.*

$$\mathbb{P}\left(Q_{CEXP} - q_s > t\right) \leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) e^{-\frac{(n(1-q)td_t)\epsilon}{2}} + e^{-\frac{td_t q^2}{2}n}$$
$$\mathbb{P}\left(Q_{GEXP} - q_s > t\right) \leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) e^{-\frac{(n(1-q)td_t)\epsilon}{2}} + e^{-\frac{td_t q^2}{2}n}$$

where $C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) \triangleq \frac{1}{4\omega} e^{\frac{\epsilon - 2\epsilon_\mu}{4}} \left(1 + e^{-\epsilon_a/2}\right)$.

the bound above reduces both additive and multiplicative constants compared to the upper bound provided in on releasing the median [AD20], and also remove

the dependence on d_t^2 which can be a very small value if we do not assume prior knowledge on the distribution.

As detailed in the proof, the term $n(1-q)td_t$ serves as a lower bound for the proportion of elements in the interval $[q_s, q_s + t]$. However, it relies on the infimum of the probability density function of the distribution, which can approach zero, limiting its practicality.

Below, we present a bound that is more interpretable and effectively highlights the influence of the data distribution, emphasizing the importance of reducing the constant factor.

Lemma 6.3. *Let $\omega > 0$ be defined as in 4. Fix $t > \omega$, $q_s = X_{\lceil qn \rceil}$ such that $\exists \eta > 0$*

$$\mathbb{E}\left(\mathbb{1}_{x \sim P \in [q_s, q_s + t]}\right) \geq \eta n.$$

Then

$$\begin{aligned} P\left(Q_{CEXP} - q_s > t\right) &\leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) \cdot e^{-\frac{(\mathbb{P}(x \sim P \in [q_s, q_s + t]) - \eta n)\epsilon}{2}} \\ &\quad + e^{-2n\eta^2} \\ P\left(Q_{GEXP} - q_s > t\right) &\leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) \cdot e^{-\frac{(\mathbb{P}(x \sim P \in [q_s, q_s + t]) - \eta n)\epsilon}{2}} \\ &\quad + e^{-2n\eta^2} \end{aligned}$$

where $C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right)$ is defined as in 6.2.

The term η is a tolerance parameter. The private estimator will produce fewer outputs in the range $[q_s, q_s + t]$ as $\mathbb{P}\left(x \sim P \in [q_s, q_s + t]\right)$ gets larger.

The lemma above shows that the theoretical guarantees of the exponential mechanism depend on the underlying distribution, if a distribution attributes little density to large regions in the output space, it gets harder to provide theoretical guarantees, This is why improving multiplicative and additive constants is important to tighten the bound in cases where the clipping endpoints are farther apart. it also shows that if the output space is large, Dividing the privacy budget between ϵ_μ, ϵ seems to be the most effective strategy to improve performance as they have the most effect on the multiplicative constant and the distribution term respectively, this is in line with our experimental findings in 8.

7 Unbiased differentially private quantile release mechanism - UBEXP

In Section 4, the mechanisms we introduced tackled both conditional bias and inefficiencies in privacy budget allocation. In this section, we delve deeper into the challenge of completely eliminating the bias. We demonstrate that fully eradicating the bias is inherently constrained by a 'no free lunch' theorem, which,

in the absence of distributional assumptions, leads to a quadratic expansion of the clipping range.

We introduce the unbiased exponential mechanism. We define $\text{Gap}_q = X^{O,'}_{(\lceil qn \rceil)} - X^{O}_{(\lceil qn \rceil)} \geq 0$

Definition 7.1 (Unbiased Exponential Mechanism - UBEXP)**.** Let $q \in (0,1)$, $b \in \mathbb{R}$, and $X^O \in [-b,b]^n$ be a dataset.

Define the minimum quantile distance:

$$mn_q = \min\left(\lceil qn \rceil - 1, n - \lceil qn \rceil\right).$$

We construct an extended dataset $X^o \in [a,b]^{2mn_q+3}$ s.t
For all $k \in [1, mn_q]$, we extend values in X^O such that

$$X^o_{(\lceil qn \rceil \pm k)} = X^O_{(\lceil qn \rceil \pm k)}.$$

and by setting boundary values:

$$X^o_{(\lceil qn \rceil - mn_q - 1)} = -b, \quad X^o_{(\lceil qn \rceil + mn_q + 1)} = b.$$

Using X^o, we define a symmetric vector X centered around the target quantile q where is new rank is $r_q = mn_q + 2$:

$$\forall k \in \{1, \ldots, mn_q + 1\} :$$

$$\begin{aligned} x_{(r_q+k)} &= x^o_{(\lceil qn \rceil)} + k\text{Gap}_q + t_{(k)}, \\ x_{(r_q-k)} &= x^o_{(\lceil qn \rceil)} - k\text{Gap}_q - t_{(k)}, \end{aligned}$$

where the term $t_{(k)}$ is given by:

$$t_{(k)} = \max\left(x^o_{(\lceil qn \rceil)} - x^o_{(\lceil qn \rceil - k)}, x^o_{(\lceil qn \rceil + k)} - x^o_{(\lceil qn \rceil)}\right).$$

This process generates a symmetric dataset X with $2 \cdot mn_q + 3$ elements within an extended range.

For interval $I_k \in \{I_1, \cdot I_{2mn_q+2}\}$, the utility function $u(I_k)$ is defined as:

$$u(I_k) = \min_{m \in \{k, k+1\}} \exp\left(-\frac{\epsilon}{4} r(m)\right) \tag{1}$$

where $r(m)$ is the rank function, defined as:

$$m \in [2mn_q + 3] \qquad r(m) = |m - r_q|$$

Here, the key components are:

- $u(I_k)$ is the utility function for interval I_k,
- $r(k)$ is the rank of the sample relative to the target quantile.

Lemma 7.2. *Let $X^O, X^{O'} \in [a,b]^n$ be two adjacent datasets, with $X^{O'}$ created by changing a value to a value larger larger than it. Let $X^o, X^{o'} \in [a,b]^{2mn_q+3}$, $X, X' \in \mathcal{R}^{2mn_q+3}$ be vectors constructed by the translations defined in Definition 7.1, $r_q = mn_q + 2$ the quantile rank under X. Define mappings $s_{l,X}, s_{r,X} : [1, mn_q] \to \mathbb{R}$ as follows:*

$$\begin{aligned} s_{l,X}(k) &= x^o_{(\lceil qn \rceil)} - x^o_{(\lceil qn \rceil - k)}, \\ s_{r,X}(k) &= x^o_{(\lceil qn \rceil + k)} - x^o_{(\lceil qn \rceil)}, \\ s_{l,X'}(k) &= x^{o'}_{(\lceil qn \rceil)} - x^{o'}_{(\lceil qn \rceil - k)}, \\ s_{r,X'}(k) &= x^{o'}_{(\lceil qn \rceil + k)} - x^{o'}_{(\lceil qn \rceil)}. \end{aligned}$$

Both mappings $s_{l,X}$ and $s_{r,X}$ are non-decreasing, with the following properties holding for all $k \in [0, mn_q]$:

$$x^o_{(\lceil qn \rceil + k)} \leq x^{o'}_{(\lceil qn \rceil + k)} \leq x^o_{(\lceil qn \rceil + k + 1)},$$

and for all $k \in [1, mn_q]$:

$$x^o_{(\lceil qn \rceil - k)} \leq x^{o'}_{(\lceil qn \rceil - k)} \leq x^o_{(\lceil qn \rceil - k + 1)}.$$

Furthermore, for any $m \geq 1$ and $j \geq 0$, the following inequalities hold:

$$x_{(r_q + j)} \geq x'_{(r_q - m)}, \quad x'_{(r_q + j)} \geq x_{(r_q - m)}.$$

Lemma 7.3. *Let $X^O, X^{O'} \in [a,b]^n$ be two adjacent datasets, with $X^{O'}$ created by changing a value to a larger value. Let $X^o, X^{o'} \in [a,b]^{2mn_q+3}, X, X' \in \mathcal{R}^{2mn_q+3}$ be vectors constructed by the translations defined in Definition 7.1, $r_q = mn_q + 2$ the quantile rank under X.*

For all $k \in [0, mn_q + 1]$, we have:

$$\begin{aligned} x_{(r_q + k)} &\geq x'_{(r_q + k - 2)}, \\ x'_{(r_q + k)} &\geq x_{(r_q + k - 2)}. \end{aligned}$$

Similarly, for all $k \in [2, mn_q + 1]$, we have:

$$\begin{aligned} x_{(r_q - k + 2)} &\geq x'_{(r_q - k)}, \\ x'_{(r_q - k + 2)} &\geq x_{(r_q - k)}. \end{aligned}$$

The above lemma shows, that the construction of X, X' from adjacent datasets makes sure that they populate the range $\mathcal{R}$ of the mechanism at a similar rate, we will use 7.3 to show that UBCEXP guarantees pure differential privacy.

Theorem 7.4 (UBCEXP DP Guarantees)**.** *7.1 satisfies pure ϵ differential privacy*

The symmetry of the intervals, also makes the overall mechanism unbiased for differentially private quantile release.

Theorem 7.5 (Bias of UBEXP)**.** *Let $X^O \in [a,b]^n$, $r_q = mn_q + 2$ the quantile rank under X. The mechanism 7.1 with the adjusted utility 1 is unbiased for differentially private median release under the swap adjacency relation.* $\mathbb{E}\left(Q_{CEXP}\left(X^O\right)\right) = x_{r_q} = x^O_{(\lceil qn \rceil)}$

8 Experiments

To ensure reproducibility and transparency, we provide an open-source implementation of our techniques at `DP_Quantiles`. The repository includes all the code necessary to replicate our experiments and apply our methods to new datasets.

The experiments section will be divided to three parts, in the first part we will show the performance of the single quantile mechanisms, then we will demonstrate how the improvements also enhance the many quantile algorithm proposed in [KSS22]. Finally, we will rely on the Gaussian sampling technique proposed in to demonstrate its effectiveness in the presence of Gaussian data. Additional information on the experimentation setup (choice of metrics, numerical stability etc) is provided in the appendix E along with the evaluation of the fully unbiased exponential mechanism in appendix E.2.

The figures below illustrate the performance of estimators designed to reduce conditional bias as a function of epsilon. The left figure depicts results for data randomly sampled from a lognormal distribution, while the right figure shows performance when a minimum distance of $Gap_q \leq x_{(\lceil qn \rceil + 1)} - x_{(\lceil qn \rceil)} \geq 100$ is enforced between the quantile and the datapoint to its right.

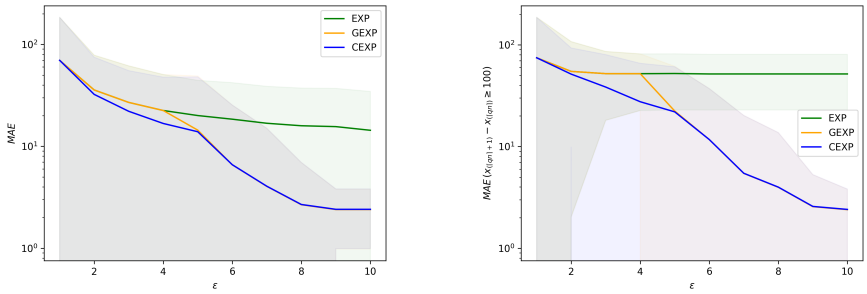

Figure 3: Results for $n = 10^3$, $X \sim \text{Lognormal}(10; 0.4)$.

We demonstrate that mechanisms allocating a portion of the privacy budget to reduce bias effectively lower the mean absolute error. However, as established in the lemma 6.3, when the interval containing the quantile is stretched, the mean absolute error of the exponential mechanism plateaus and ceases to improve further with increasing epsilon (these distributions can exist in real-world

datasets).

Below we demonstrate the improvement of the many quantile algorithm introduced by [KSS22], by changing the base single quantile algorithm to reduce conditional bias.

Distribution	Mechanism	MAE
Uniform $\left[0, 4 \times 10^4\right]$	**CEXP**	$\mathbf{4,53 \times 10^2 \pm 1,81 \times 10^1}$
	GEXP	$4,56 \times 10^2 \pm 1,64 \times 10^1$
	EXP	$4,56 \times 10^2 \pm 1,64 \times 10^1$
Lognormal $[10, 0.4]$	**CEXP**	$\mathbf{5,56 \times 10^2 \pm 2,77 \times 10^1}$
	GEXP	$5,65 \times 10^2 \pm 2,74 \times 10^1$
	EXP	$5,65 \times 10^2 \pm 2,74 \times 10^1$
Normal $\left(3 \times 10^4, (8 \times 10^3)^2\right)$	**CEXP**	$\mathbf{6,74 \times 10^2 \pm 2,58 \times 10^1}$
	GEXP	$6,82 \times 10^2 \pm 2,31 \times 10^1$
	EXP	$6,82 \times 10^2 \pm 2,31 \times 10^1$

Table 1: Many quantiles mechanism - $q = (0,2; 0,8)$ - $\epsilon = 5$ - $n = 4 \times 10^2$ - $[a, b] = [-10^4, 10^4]$

8.1 Application to gaussian perturbed data

Using a Gaussian distribution instead of the usual uniform distribution in the exponential mechanism can be advantageous when data is randomized. This approach allows the mechanism to select the highest utility interval and return an already perturbed sample without additional randomization. the randomized data quantile mechanism trades pure differential privacy for approximate differential privacy, offering greater utility at the cost of slightly relaxed privacy guarantees. Decreasing δ increases the variance of the estimator

Distribution	Mechanism	MAE
Uniform $[0, 4 \times 10^4]$	**EXP**	$65,16 \pm 69,13$
	RQM	$\mathbf{54,50 \pm 89,88}$
Lognormal $[10, 0.4]$	**EXP**	$50,26 \pm 64,15$
	RQM	$\mathbf{37,80 \pm 73,30}$
Normal$\left(3 \times 10^4, (8 \times 10^3)^2\right)$	**EXP**	$42,82 \pm 46,91$
	RQM	$\mathbf{32,54 \pm 63,67}$

Table 2: MAE values for different distributions and mechanisms - $\sigma = 20$ - Releasing the quantile $q \sim \text{Unif}\,(0, 1; 0, 9)$ - $\epsilon = 6$ - $\delta = 5 \times 10^{-2}$ - $D = 0$ - $n = 4 \times 10^2$ - $[a, b] = [-10^4, 10^4]$

9 Conclusion

In this paper, we developed novel mechanisms that effectively reduce the conditional bias in the existing differentially private quantile algorithms that rely on the exponential mechanism, leading to a more flexible attribution of the privacy budget. Our mechanisms outperform the current state-of-the-art (SOTA) in both single and many quantile settings, demonstrating significant improvements in privacy and utility. The fully unbiased mechanism, while achieving comparable performance to the SOTA in the presence of distributional assumptions, highlights the trade-offs involved in maintaining unbiasedness while avoiding distributional assumptions.

We also successfully adapted the exponential mechanism to incorporate for gaussian data, using it as a case study to show the potential for leveraging Gaussian data to improve privacy guarantees.

Looking forward, future work can focus on enhancing the fully unbiased mechanism to improve its practicality.

References

[AD20] Hilal Asi and John C Duchi. Near instance-optimality in differential privacy. *arXiv preprint arXiv:2005.10630*, 2020.

[AMS+20] Daniel Alabi, Audra McMillan, Jayshree Sarathy, Adam Smith, and Salil Vadhan. Differentially private simple linear regression. *arXiv preprint arXiv:2007.05157*, 2020.

[App17] Differential Privacy Team Apple. Learning with privacy at scale, 2017.

[CSS11] T-H Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):1–24, 2011.

[DGHM+22] Jörg Drechsler, Ira Globus-Harris, Audra Mcmillan, Jayshree Sarathy, and Adam Smith. Nonparametric differentially private confidence intervals for the median. *Journal of Survey Statistics and Methodology*, 10(3):804–829, 2022.

[DKY17] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30, 2017.

[DL09] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 371–380, 2009.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.

[DR+14] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[DRV10] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st annual symposium on foundations of computer science*, pages 51–60. IEEE, 2010.

[EPK14] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.

[GJK21] Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. Differentially private quantiles. In *International Conference on Machine Learning*, pages 3713–3722. PMLR, 2021.

[Hoe94] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.

[JLY+24] Yangdi Jiang, Yi Liu, Xiaodong Yan, Anne-Sophie Charest, Linglong Kong, and Bei Jiang. Analysis of differentially private synthetic data: A measurement error approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 21206–21213, 2024.

[KSS22] Haim Kaplan, Shachar Schnapp, and Uri Stemmer. Differentially private approximate quantiles. In *International Conference on Machine Learning*, pages 10751–10761. PMLR, 2022.

[LGG23a] Clément Lalanne, Aurélien Garivier, and Rémi Gribonval. Private statistical estimation of many quantiles. In *International Conference on Machine Learning*, pages 18399–18418. PMLR, 2023.

[LGG+23b] Clément Lalanne, Clément Gastaud, Nicolas Grislain, Aurélien Garivier, and Rémi Gribonval. Private quantiles estimation in the presence of atoms. *Information and Inference: A Journal of the IMA*, 12(3):2197–2223, 2023.

[MG20] Andrés Muñoz Medina and Jenny Gillenwater. Duff: A dataset-distance-based utility function family for the exponential mechanism. *arXiv preprint arXiv:2010.04235*, 2020.

[MSI23] Cory McCartan, Tyler Simko, and Kosuke Imai. Making differential privacy work for census data users. *Harvard Data Science Review*, 5(4), November 2023.

[MT07] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

[NRS07] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.

[Smi11] Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 813–822, 2011.

[Swe02] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05):557–570, 2002.

Appendix

A Centered simple differentially private quantile release mechanisms

A.1 Centered Exponential Mechanism - CEXP

A.1.1 Proof of Lemma 4.3

Lemma. Let $X \in \mathcal{R}_X$ constructed following 4

for all $k \in [0, n - r_q + 1]$

$$x_{(r_q+k)} \leq x'_{(r_q+k+1)}$$
$$x'_{(r_q+k)} \leq x_{(r_q+k+1)}$$

for all $k \in [0, r_q - 2]$

$$x_{(r_q-k-1)} \leq x'_{(r_q-k)}$$
$$x'_{(r_q-k-1)} \leq x_{(r_q-k)}$$

Proof. For convenience we will refer to the new rank of the quantile as $\lceil qn \rceil$

Let $X, X' \in \mathcal{R}$ be constructed from adjacent datasets as defined in 4.1. We consider an adjacent dataset created by changing a value to a larger value. (the other is similar, it suffices to exchange X, X')

One can see that in this case the following inequalities hold.

$$\forall k \in [0, n - \lceil qn \rceil] \quad x^o_{\lceil qn \rceil + k} \leq x_{\lceil qn \rceil + k}$$
$$\forall k \in [0, \lceil qn \rceil - 1] \quad x^o_{\lceil qn \rceil - k} \geq x_{\lceil qn \rceil - k}$$

The above holds because we stretch the samples to ensure a distance of at least 2ω between consecutive datapoints. We also have

$$\forall k \in [1, n + 2] \quad x^o_k \leq x^{o,'}_k$$
$$\forall k \in [1, n + 1] \quad x^{o,'}_k \leq x^o_{k+1}$$
$$\forall k \in [1, n + 1] \quad x_k + 2\omega \leq x_{k+1}$$
$$\forall k \in [1, n + 1] \quad x'_k + 2\omega \leq x'_{k+1}$$

The second inequality above holds because Synthetic datapoints (a,b) are added to the datasets in the preprocessing phase.

To prove the four inequalities we will proceed by induction

we will start by showing that

for all $k \in [0, n - \lceil qn \rceil + 1]$

$$x_{(\lceil qn \rceil + k)} \leq x'_{(\lceil qn \rceil + k + 1)}$$
$$x'_{(\lceil qn \rceil + k)} \leq x_{(\lceil qn \rceil + k + 1)}$$

for $k = 0$

$$x'_{(\lceil qn \rceil + 1)} \geq x'_{(\lceil qn \rceil)} \geq x^o_{(\lceil qn \rceil)} = x_{(\lceil qn \rceil)}$$
$$x_{(\lceil qn \rceil + 1)} \geq x^o_{(\lceil qn \rceil + 1)} \geq x^{o,'}_{(\lceil qn \rceil)} = x'_{(\lceil qn \rceil)}$$

Fix $k \in \mathbb{N}^*$, we assume that both inequalities hold for all k in the range $[0, k-1]$.

For k:

$$x_{(\lceil qn \rceil + k)} = x_{(\lceil qn \rceil + k-1)} + \max\left(2\omega, x^o_{(\lceil qn \rceil + k)} - x_{(\lceil qn \rceil + k-1)}\right).$$

If $\max\left(2\omega, x^o_{(\lceil qn \rceil + k)} - x_{(\lceil qn \rceil + k-1)}\right) = x^o_{(\lceil qn \rceil + k)} - x_{(\lceil qn \rceil + k-1)}$, then:

$$x_{(\lceil qn \rceil + k)} = x^o_{(\lceil qn \rceil + k)} \leq x^{o,'}_{(\lceil qn \rceil + k)} \leq x'_{(\lceil qn \rceil + k+1)}.$$

If $\max\left(2\omega, x^o_{(\lceil qn \rceil + k+1)} - x_{(\lceil qn \rceil + k)}\right) = 2\omega$, then:

$$\begin{aligned} x_{(\lceil qn \rceil + k)} &\leq x_{(\lceil qn \rceil + k-1)} + 2\omega, \\ x_{(\lceil qn \rceil + k)} &\leq x'_{(\lceil qn \rceil + k)} + 2\omega, \\ x_{(\lceil qn \rceil + k)} &\leq x'_{(\lceil qn \rceil + k+1)}. \end{aligned}$$

Similarly, we consider:

$$x'_{(\lceil qn \rceil + k)} \leq x'_{(\lceil qn \rceil + k-1)} + \max\left(2\omega, x^{o,'}_{(\lceil qn \rceil + k)} - x'_{(\lceil qn \rceil + k-1)}\right).$$

If $\max\left(2\omega, x^{o,'}_{(\lceil qn \rceil + k)} - x'_{(\lceil qn \rceil + k-1)}\right) = x^{o,'}_{(\lceil qn \rceil + k)} - x'_{(\lceil qn \rceil + k-1)}$, then:

$$x'_{(\lceil qn \rceil + k)} = x^{o,'}_{(\lceil qn \rceil + k)} \leq x^o_{(\lceil qn \rceil + k+1)} \leq x_{(\lceil qn \rceil + k+1)}.$$

If $\max\left(2\omega, x^{o,'}_{(\lceil qn \rceil + k+1)} - x'_{(\lceil qn \rceil + k)}\right) = 2\omega$, then:

$$\begin{aligned} x'_{(\lceil qn \rceil + k)} &= x'_{(\lceil qn \rceil + k-1)} + 2\omega, \\ x'_{(\lceil qn \rceil + k)} &\leq x_{(\lceil qn \rceil + k)} + 2\omega, \\ x'_{(\lceil qn \rceil + k)} &\leq x_{(\lceil qn \rceil + k+1)}. \end{aligned}$$

Now we will show inequalities for the left side of the quantile value, we will show that

for all $k \in [0, \lceil qn \rceil - 2]$

$$\begin{aligned} x_{(\lceil qn \rceil - k-1)} &\leq x'_{(\lceil qn \rceil - k)} \\ x'_{(\lceil qn \rceil - k-1)} &\leq x_{(\lceil qn \rceil - k)} \end{aligned}$$

Both inequalities holds for $k = 0$

Fix $k \geq 0$ and suppose that both inequalities hold for $\{0, .., k\}$ we will show that the inequalities hold for $k+1$

$$x_{(\lceil qn \rceil - k)} = x_{(\lceil qn \rceil - k+1)} - \max\left(2\omega, x_{(\lceil qn \rceil - k+1)} - x^o_{(\lceil qn \rceil - k)}\right)$$

if the maximum value is equal to $x_{(\lceil qn \rceil - k+1)} - x^o_{(\lceil qn \rceil - k)}$ we will have

$$\begin{aligned} x_{(\lceil qn \rceil - k)} &= x^o_{(\lceil qn \rceil - k)} \\ x_{(\lceil qn \rceil - k)} &\geq x^{o,'}_{(\lceil qn \rceil - k-1)} \\ x_{(\lceil qn \rceil - k)} &\geq x'_{(\lceil qn \rceil - k-1)} \end{aligned}$$

if the maximum value is equal to 2ω we will have

$$\begin{aligned} x_{(\lceil qn \rceil - k)} &= x_{(\lceil qn \rceil - k+1)} - 2\omega \\ x_{(\lceil qn \rceil - k)} &\geq x'_{(\lceil qn \rceil - k)} - 2\omega \\ x_{(\lceil qn \rceil - k)} &\geq x'_{(\lceil qn \rceil - k-1)} \end{aligned}$$

Above we used the second inequality in the induction hypotheses

Similarly we have

$$x'_{(\lceil qn \rceil - k)} = x'_{(\lceil qn \rceil - k+1)} - \max\left(2\omega, x'_{(\lceil qn \rceil - k+1)} - x^{o,'}_{(\lceil qn \rceil - k)}\right)$$

if the maximum value is equal to $x_{(\lceil qn \rceil - k+1)} - x^o_{(\lceil qn \rceil - k)}$ we will have

$$\begin{aligned} x'_{(\lceil qn \rceil - k)} &= x^{o,'}_{(\lceil qn \rceil - k)} \\ x'_{(\lceil qn \rceil - k)} &\geq x^o_{(\lceil qn \rceil - k)} \\ x'_{(\lceil qn \rceil - k)} &\geq x_{(\lceil qn \rceil - k-1)} \end{aligned}$$

if the maximum value is equal to 2ω we will have

$$\begin{aligned} x'_{(\lceil qn \rceil - k)} &= x'_{(\lceil qn \rceil - k+1)} - 2\omega \\ x'_{(\lceil qn \rceil - k)} &\geq x_{(\lceil qn \rceil - k)} - 2\omega \\ x'_{(\lceil qn \rceil - k)} &\geq x_{(\lceil qn \rceil - k-1)} \end{aligned}$$

□

A.1.2 Proof of Theorem 4.4

Theorem (Centered exponential mechanism DP guarantees). Sampling an element in the interval $\mathcal{R}_X$ with the density defined in 4.1 satisfies pure $\epsilon_\mu + \epsilon_a + \epsilon$ differential privacy.

Proof. Let $X, X' \in \mathcal{R}$ be constructed from adjacent datasets as defined in 4.1. We consider an adjacent dataset created a smaller value to a larger value (it suffices to exchange X, X' to get the second case)

To prove differential privacy, we can consider an output $o \in [x_{(\lceil qn \rceil + k)}, x_{(\lceil qn \rceil + k + 1)}]$ for $k \in \mathbb{Z}$.

We showed in 4.3 that (by convention, values outside the boundaries will be either equal to a or b):

$$x'_{(\lceil qn \rceil + k - 1)} \leq x_{(\lceil qn \rceil + k)} \leq o \leq x_{(\lceil qn \rceil + k + 1)} \leq x'_{(\lceil qn \rceil + k + 2)}.$$

This means that if $o \in I_{2p-1} = [x_{(p)} - 2\omega, x_{(p)} + 2\omega]$, then we have:

$$x'_{(p-1)} + 2\omega \leq o \leq x'_{(p+1)} + 2\omega,$$

therefore $o \in \{I'_{2p-1}, I'_{2p-2}, I'_{2p}\}$, which means that the rank of the interval it falls in, in the ranges created by adjacent datasets, changes by at most two.

Now, if $o \in I_{2p} = [x_{(p)} + 2\omega, x_{(p+1)} - 2\omega]$, we will have:

$$x'_{(p-1)} + 2\omega \leq x_{(p)} + 2\omega \leq o \leq x_{(p+1)} - 2\omega \leq x'_{(p+1)} - 2\omega,$$

leading to $o \in \{I'_{2p-1}, I'_{2p-2}, I'_{2p}\}$.

First, we note that the events $\{E_k\}_{k \in [2n+3]}$ are disjoint by definition, as we only sample from one interval. The utility function is also injective on the range $\mathcal{R}$.

Both distributions Q^X_{CEXP} and $Q^{X'}_{CEXP}$ have support in $\mathcal{R}$. For any possible output $o \in \mathcal{R}$ by the centered exponential mechanism, we have:

$$\begin{aligned} \frac{f(o \mid X)}{f(o \mid X')} &= \frac{\sum_{k \in [2n+3]} f(o \mid E_k) \mathbb{P}(E_k)}{\sum_{k \in [2n+3]} f(o \mid E'_k) \mathbb{P}(E'_k)} \\ &= \frac{\sum_{I_k \in \mathcal{R}_X} \mathbb{1}_{o \in I_k} u(I_k)}{\sum_{I'_k \in \mathcal{R}_{X'}} \mathbb{1}_{o \in I'_k} u(I'_k)} \times R_{x'|x}, \end{aligned}$$

where

$$R_{x'|x} \triangleq \frac{\sum_{I_{2k+1} \in \mathcal{R}_X} 2\omega u(I_{2k+1}) + \sum_{I_{2k} \in \mathcal{R}_X} 2(x_{(k+1)} - x_{(k)} - 2\omega) u(I_{2k})}{\sum_{I_{2k+1} \in \mathcal{R}_{X'}} 2\omega u(I'_{2k+1}) + \sum_{I_{2k} \in \mathcal{R}_{X'}} 2(x'_{(k+1)} - x'_{(k)} - 2\omega) u(I'_{2k})}.$$

We define the step function on the entire range $u : \mathcal{R} \to \mathbb{R}^+$. We denote the union $U = \left(I'_{2\lceil qn \rceil - 1} \cup I_{2\lceil qn \rceil - 1}\right)$.

$$\forall o \in \mathcal{R} \setminus U \quad u(o, X') \leq e^{\frac{\epsilon + \epsilon_a}{2}} u(o, X).$$

Using Lemma 4.3, we can conclude that if $u(o, X') = e^{\epsilon_\mu}$ within the interval $I'_{2\lceil qn \rceil - 1}$, then $u(o, X')$ must be either $e^{-\epsilon}$ or $e^{-\epsilon - \frac{\epsilon_a}{2}}$ on $U \setminus I'_{2\lceil qn \rceil - 1}$, while satisfying the inequality.

Formally, there exists a subdivision $\{d\}_{i \in [r]}$ of $[a, b]$ and an index $j \in [r]$ such that $[d_j, d_{j+1}] \cup [d_{j+2}, d_{j+3}] \subseteq I_{2\lceil qn \rceil - 1}$ and $[d_{j+1}, d_{j+2}] = I'_{2\lceil qn \rceil - 1}$.

$$u(o, X') \leq \max_{o \in [d_j, d_{j+1}] \cup [d_{j+2}, d_{j+3}]} u(o, X') \leq e^{-\epsilon}.$$

Additionally, we have the following ϵ-indistinguishability within the same dataset due to the closeness in ranks:

$$\forall o \in I'_{2\lceil qn \rceil - 1}, \quad u(I_{2\lceil qn \rceil - 1}, X) \leq e^{\frac{\epsilon + \epsilon_a}{2} + \epsilon_\mu} u(o, X).$$

Then, we can express the ratio as follows:

$$R_{x'|x} = \frac{2\omega e^{\epsilon_\mu} + \int_{d_j}^{d_{j+1}} u(o, X') + \int_{d_{j+2}}^{d_{j+3}} u(o, X')\, do + \sum_{\mathcal{R} - (I'_{2\lceil qn \rceil - 1} \cup I_{2\lceil qn \rceil - 1})} \int u(o, X')\, do}{2\omega e^{\epsilon_\mu} + \sum_{\mathcal{R} - I_{2\lceil qn \rceil - 1}} \int u(o, X)\, do},$$

$$R_{x'|x} \leq \frac{2\omega e^{\epsilon_\mu} + e^{-\epsilon - \epsilon_\mu} \int_{I'_{2\lceil qn \rceil - 1}} e^{\epsilon_\mu}\, do + \sum_{\mathcal{R} - (I'_{2\lceil qn \rceil - 1} \cup I_{2\lceil qn \rceil - 1})} \int u(o, X')\, do}{2\omega e^{\epsilon_\mu} + \sum_{\mathcal{R} - I_{2\lceil qn \rceil - 1}} \int u(o, X)\, do},$$

$$R_{x'|x} \leq \frac{\int_{b_1}^{b_2} u(o, X')\, do + e^{-\epsilon - \epsilon_\mu} \int_{I'_{2\lceil qn \rceil - 1}} u(I_{2\lceil qn \rceil - 1}, X)\, do + \sum_{\mathcal{R} - (I'_{2\lceil qn \rceil - 1} \cup I_{2\lceil qn \rceil - 1})} \int u(o, X')\, do}{2\omega e^{\epsilon_\mu} + \sum_{\mathcal{R} - I_{2\lceil qn \rceil - 1}} \int u(o, X)\, do}.$$

Above, we used that $u(o, I_{2\lceil qn \rceil - 1}) = e^{\epsilon_\mu}$ and that $|[d_j, d_{j+1}] \cup [d_{j+2}, d_{j+3}]| \leq \left| I'_{2\lceil qn \rceil - 1} \right|$. Furthermore, $u(o, X') \leq \max_{o \in [d_j, d_{j+1}] \cup [d_{j+2}, d_{j+3}]} u(o, X') \leq e^{-\epsilon}$. To continue, it suffices to use the ϵ-indistinguishability for intervals of the same rank under adjacent datasets.

That is:

$$\forall o \in I'_{2\lceil qn \rceil - 1}, \quad u(I_{2\lceil qn \rceil - 1}, X) \leq e^{\frac{\epsilon + \epsilon_a}{2} + \epsilon_\mu} u(o, X).$$

$$R_{x'|x} \leq \frac{2\omega e^{\epsilon_\mu} + e^{\frac{\epsilon + \epsilon_a}{2}} \int_{I'_{2\lceil qn \rceil - 1}} u(o, X)\, do + \sum_{\mathcal{R} - (I'_{2\lceil qn \rceil - 1} \cup I_{2\lceil qn \rceil - 1})} \int u(o, X')\, do}{2\omega e^{\epsilon_\mu} + \sum_{\mathcal{R} - I_{2\lceil qn \rceil - 1}} \int u(o, X)\, do},$$

$$R_{x'|x} \leq e^{\frac{\epsilon + \epsilon_a}{2}} \cdot \frac{2\omega e^{\epsilon_\mu} + \int_{I'_{2\lceil qn \rceil - 1}} u(o, X)\, do + \sum_{\mathcal{R} - (I'_{2\lceil qn \rceil - 1} \cup I_{2\lceil qn \rceil - 1})} \int u(o, X)\, do}{2\omega e^{\epsilon_\mu} + \sum_{\mathcal{R} - I_{2\lceil qn \rceil - 1}} \int u(o, X)\, do},$$

$$R_{x'|x} \leq e^{\frac{\epsilon + \epsilon_a}{2}}.$$

Now consider an arbitrary output in $\mathcal{R}$. The ranges complete the interval $\mathcal{R}$ and every possible part is covered once. This means that we can write the ratio of densities as a function of $m, j \in [2n + 3]$ that denote the rank of interval covering o in $\mathcal{R}_X, \mathcal{R}_{X'}$:

$$\begin{aligned} \frac{f(o \mid X)}{f(o \mid X')} &\leq e^{\frac{\epsilon + \epsilon_a}{2}} \frac{\sum_{I_k \in \mathcal{R}_X} \mathbb{1}_{o \in I_k} u(I_k)}{\sum_{I'_k \in \mathcal{R}_{X'}} \mathbb{1}_{o \in I'_k} u(I'_k)}, \\ &\leq e^{\frac{\epsilon + \epsilon_a}{2}} \frac{u(I_m)}{u(I'_j)}. \end{aligned}$$

Since the same output can move from an odd rank interval to an even rank interval, or from an odd rank maximum utility interval to an even rank interval, in the worst case we will have:

$$\frac{u(I_m)}{u(I'_j)} \leq e^{\frac{\epsilon_q}{2}+\epsilon_\mu+\frac{\epsilon}{2}}.$$

The above holds since $|m-j| \leq 2$, which gives us:

$$\frac{f(o \mid X)}{f(o \mid X')} \leq e^{\epsilon+\epsilon_\mu+\epsilon_a}.$$

Therefore, the mechanism Q_{CEXP} is $(\epsilon+\epsilon_\mu+\epsilon_a)$-differentially private. □

A.2 Gaussian Exponential mechanism - GEXP

For convenience we use $\lceil qn \rceil$ instead of r_q to denote the quantity we want to estimate

Definition A.1 (Gaussian Exponential Mechanism)**.** Let $q \in (0,1)$, $a, b \in \mathbb{R}$, $\omega, \sigma_{GEXP} > 0$, and $X^O \in [a,b]^n$ be a dataset.

We construct an extended dataset $X^o \in [a,b]^{n+2}$ by adding boundary values:

$$X^o_{(1)} = a, \quad X^o_{(n+2)} = b.$$

Next, we create a new vector $X \in \mathcal{R} = [a-(\lceil qn \rceil -1)\omega, b+(n-\lceil qn \rceil -1)\omega]^{n+2}$ by ensuring a distance of at least $2\omega > 0$ between samples.

Formally, for $k \geq 1$, we define:

$$\begin{aligned} X_{(\lceil qn \rceil +k)} &= X_{(\lceil qn \rceil +k-1)} + t^+(k), \\ X_{(\lceil qn \rceil -k)} &= X_{(\lceil qn \rceil -k+1)} + t^-(k). \end{aligned}$$

where $t^+(k) = \max\left(2\omega, X^o_{(\lceil qn \rceil +k)} - X_{(\lceil qn \rceil +k-1)}\right)$ and $t^-(k) = -\max\left(2\omega, X_{(\lceil qn \rceil -k+1)} - X^o_{(\lceil qn \rceil -k)}\right)$

The dataset divides $\mathcal{R}$ into $2n+3$ intervals, denoted as I_k, defined as follows:

$$I_k = \begin{cases} [x_{(k)} - \omega, x_{(k)} + \omega], & \text{if } k \text{ is odd,} \\ [x_{(k)} + \omega, x_{(k)} - \omega], & \text{if } k \text{ is even.} \end{cases}$$

Let E_m denote the event of selecting an interval I_m from the range $\mathcal{R}_X$.

The mechanism generates an output $o \in [a-(\lceil qn \rceil -1)\omega, b+(n-\lceil qn \rceil -1)\omega]$ with the following probability density:

$$f(o) = \sum_{I_k \in \mathcal{R}_X} f_{I_k}(o) \cdot \mathbb{P}(E_k),$$

where the density function $f_{I_k}(o)$ is determined by the interval I_k:

$$f_{I_k}(o) = \begin{cases} f_{\mathcal{TN}\left(x_k, x_k-\omega, x_k+\omega, \sigma^2_{GEXP}\right)} \mathbb{1}_{I_k}(o), & \text{if } k \text{ is odd,} \\ \frac{1}{x_{(k+1)}-x_{(k)}-2\omega} \mathbb{1}_{I_k}(o), & \text{if } k \text{ is even.} \end{cases}$$

The probability of selecting interval E_k is given by:

$$\mathbb{P}(E_k) \propto \begin{cases} (x_{(k+1)} - x_{(k)} - 2\omega) \cdot u(I_k), & \text{if } k \text{ is even,} \\ 2\omega \cdot u(I_k), & \text{if } k \text{ is odd.} \end{cases}$$

The utility function $u(I_k)$ is defined as:

$$u(I_k) = \begin{cases} \exp(\epsilon_\mu) \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is odd and } r(k) = 0, \\ \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is odd and } r(k) \neq 0, \\ \exp\left(-\frac{\epsilon_a}{2}\right) \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is even.} \end{cases}$$

Here, $r(k)$ is the rank function, defined as:

$$r(k) = \left| \left\lceil \frac{k}{2} \right\rceil - \lceil q \cdot n \rceil \right|,$$

where:

- $u(I_k)$ is the utility function for interval I_k,
- q is the target quantile, and
- $r(k)$ is the rank of the sampled interval relative to the target quantile.

A.2.1 Proof of Theorem 4.7

Theorem (Gaussian Exponential Mechanism DP Guarantees). Let $\sigma_{GEXP} > 0$. Sampling an element in the range $\mathcal{R}_X \subseteq [a - (\lceil qn \rceil - 1)\omega, b + (n - \lceil qn \rceil - 1)\omega]$ with the density defined in Definition 4.6 satisfies pure differential privacy with

$$\max \begin{pmatrix} \epsilon_a + \epsilon_\mu + \epsilon + \log(\frac{2\omega}{W}) \\ \frac{\omega^2}{2\sigma^2_{GEXP}} + \epsilon_\mu + \frac{\epsilon_a}{2} + \epsilon \\ \frac{\omega^2}{2\sigma^2_{GEXP}} + \epsilon + \log(\frac{W}{2\omega}) \\ \epsilon_a + \epsilon \end{pmatrix}$$

where $W = \sqrt{2\pi}\sigma_{GEXP}\left(\Phi\left(\frac{\omega}{\sigma_{GEXP}}\right) - \Phi\left(-\frac{\omega}{\sigma_{GEXP}}\right)\right)$.

Here, Φ denotes the cumulative distribution function (CDF) of a standard normal distribution.

Proof. Let $X, X' \in \mathcal{R}$ be constructed from adjacent datasets as defined in 4.1. We consider an adjacent dataset created a smaller value to a larger value (it suffices to exchange X, X' to get the second case)

We proceed similar to the proof

$$\frac{f(o \mid X)}{f(o \mid X')} = \frac{\sum_{k\in[2n+3]} f(o \mid E_k)\mathbb{P}(E_k)}{\sum_{k\in[2n+3]} f(o \mid E'_k)\mathbb{P}(E'_k)},$$

As shown in A.2.1 for all adjacent datasets X, X', and for all outputs in $\mathcal{R}$, we have $\left|\lceil \frac{k'}{2} \rceil - \lceil \frac{k}{2} \rceil\right| \leq 1$.

We use the same trick to bound the ratio of integrals, the pmf for interval selecting is exactly the same as CEXP

$$R_{x'|x} \leq e^{\frac{\epsilon+\epsilon_a}{2}}.$$

Now consider an arbitrary output in $\mathcal{R}$. The ranges complete the interval $\mathcal{R}$ and every possible part is covered once. This means that we can write the ratio of densities for $m, j \in [2n+3]$ that denote the rank of interval covering o in $\mathcal{R}_X, \mathcal{R}_{X'}$,

we have two kinds of intervals and we will have four cases in total, we take the worst-case upper bound for each of them

$$\frac{f(o \mid X, E_m)\mathbb{P}(E_m)}{f(o \mid X, E'_j)\mathbb{P}(E'_j)} \leq \begin{cases} \frac{2\omega}{W} e^{\frac{\epsilon_a}{2}+\epsilon_\mu+\frac{\epsilon}{2}}, & \text{if } o \text{ falls in an odd rank interval in } \mathcal{R}_X \\ & \text{and an even rank interval in } \mathcal{R}_{X'}, \\ e^{\frac{\omega^2}{2\sigma^2_{GEXP}}+\epsilon_\mu+\frac{\epsilon}{2}}, & \text{if } o \text{ falls in an odd rank interval in both cases,} \\ \frac{W}{2\omega} e^{\frac{\omega^2}{2\sigma^2_{GEXP}}-\frac{\epsilon_a}{2}+\frac{\epsilon}{2}}, & \text{if } o \text{ falls in an odd rank interval in } \mathcal{R}_X \\ & \text{and an even rank interval in } \mathcal{R}_{X'}, \\ e^{\frac{\epsilon_a}{2}+\frac{\epsilon}{2}}, & \text{if } o \text{ falls in an even rank interval in both } \mathcal{R}_X \text{ and } \mathcal{R}_{X'}. \end{cases}$$

The above holds since $|m - j| \leq 2$, which gives us:

$$\frac{f(o \mid X)}{f(o \mid X')} \leq \max\left(e^{\epsilon_a+\epsilon_\mu+\epsilon+\log(\frac{2\omega}{W})}, e^{\frac{\omega^2}{2\sigma^2_{GEXP}}+\epsilon_\mu+\frac{\epsilon_a}{2}+\epsilon}, e^{\frac{\omega^2}{2\sigma^2_{GEXP}}+\epsilon+\log(\frac{W}{2\omega})}, e^{\epsilon_a+\epsilon}\right)$$

Therefore, the mechanism Q_{GEXP} is

$$\max\begin{pmatrix} \epsilon_a + \epsilon_\mu + \epsilon + \log(\frac{2\omega}{W}) \\ \frac{\omega^2}{2\sigma^2_{GEXP}} + \epsilon_\mu + \frac{\epsilon_a}{2} + \epsilon \\ \frac{\omega^2}{2\sigma^2_{GEXP}} + \epsilon + \log(\frac{W}{2\omega}) \\ \epsilon_a + \epsilon \end{pmatrix}$$

-differentially private.

□

A.3 Laplace Exponential mechanism - LEXP

Definition A.2 (Laplace Exponential Mechanism)**.** Let $q \in (0,1)$, $a, b \in \mathbb{R}$, $\omega, b_{LEXP} > 0$, and $X^O \in [a,b]^n$ be a dataset.

We construct an extended dataset $X^o \in [a,b]^{n+2}$ by adding boundary values:

$$X^o_{(1)} = a, \quad X^o_{(n+2)} = b.$$

Next, we create a new vector $X \in \mathcal{R} = [a-(\lceil qn \rceil -1)\omega, b+(n-\lceil qn \rceil -1)\omega]^{n+2}$ by ensuring a distance of at least $2\omega > 0$ between samples.

Formally, for $k \geq 1$, we define:

$$\begin{aligned} X_{(r_q+k)} &= X_{(r_q+k-1)} + t^+(k), \\ X_{(r_q-k)} &= X_{(r_q-k+1)} + t^-(k). \end{aligned}$$

where $t^+(k) = \max\left(2\omega, X^o_{(\lceil qn \rceil +k)} - X_{(r_q+k-1)}\right)$ and $t^-(k) = -\max\left(2\omega, X_{(r_q-k+1)} - X^o_{(\lceil qn \rceil -k)}\right)$

The dataset divides $\mathcal{R}$ into $2n+3$ intervals, denoted as I_k, defined as follows:

$$I_k = \begin{cases} [x_{(k)} - \omega, x_{(k)} + \omega], & \text{if } k \text{ is odd,} \\ [x_{(k)} + \omega, x_{(k)} - \omega], & \text{if } k \text{ is even.} \end{cases}$$

Let E_m denote the event of selecting an interval I_m from the range $\mathcal{R}_X$.

The mechanism generates an output $o \in [a-(\lceil qn \rceil -1)\omega, b+(n-\lceil qn \rceil -1)\omega]$ with the following probability density:

$$f(o) = \sum_{I_k \in \mathcal{R}_X} f_{I_k}(o) \cdot \mathbb{P}(E_k),$$

where the density function $f_{I_k}(o)$ is determined by the interval I_k:

$$f_{I_k}(o) = \begin{cases} f_{\mathrm{TLap}(x_k, x_k-\omega, x_k+\omega, b_{LEXP})} \mathbb{1}_{I_k}(o), & \text{if } k \text{ is odd,} \\ \frac{1}{x_{(k+1)} - x_{(k)} - 2\omega} \mathbb{1}_{I_k}(o), & \text{if } k \text{ is even.} \end{cases}$$

The probability of selecting interval E_k is given by:

$$\mathbb{P}(E_k) \propto \begin{cases} (x_{(k+1)} - x_{(k)} - 2\omega) \cdot u(I_k), & \text{if } k \text{ is even,} \\ 2\omega \cdot u(I_k), & \text{if } k \text{ is odd.} \end{cases}$$

The utility function $u(I_k)$ is defined as:

$$u(I_k) = \begin{cases} \exp(\epsilon_\mu) \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is odd and } r(k) = 0, \\ \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is odd and } r(k) \neq 0, \\ \exp\left(-\frac{\epsilon_a}{2}\right) \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is even.} \end{cases}$$

Here, $r(k)$ is the rank function, defined as:

$$r(k) = \left| \left\lceil \frac{k}{2} \right\rceil - r_q \right|,$$

where:

- $u(I_k)$ is the utility function for interval I_k,
- q is the target quantile, and
- $r(k)$ is the rank of the sampled interval relative to the target quantile.

A.3.1 Privacy guarantees

Theorem (Laplace Exponential Mechanism DP Guarantees). Let $b_{LEXP} > 0$. Sampling an element in the range $\mathcal{R}_X \subseteq [a - (\lceil qn \rceil - 1)\omega, b + (n - \lceil qn \rceil - 1)\omega]$ with the density defined in Definition A.2 satisfies pure differential privacy with

$$\max \begin{pmatrix} \epsilon_a + \epsilon_\mu + \epsilon + \log(\frac{2\omega}{W}) \\ \frac{\omega}{b_{LEXP}} + \epsilon_\mu + \frac{\epsilon_a}{2} + \epsilon \\ \frac{\omega}{b_{LEXP}} + \epsilon + \log(\frac{W}{2\omega}) \\ \epsilon_a + \epsilon \end{pmatrix}$$

where $W = 2b_{LEXP}\left(F\left(\frac{\omega}{b_{LEXP}}\right) - F\left(-\frac{\omega}{b_{LEXP}}\right)\right)$.

Here, F denotes the cumulative distribution function (CDF) of a standard Laplace distribution.

Proof. Let $X, X' \in \mathcal{R}$ be constructed from adjacent datasets as defined in 4.1.

The proof is simiar to A.2.1, we just need to bound the truncated laplace distribtution instead of the truncated normal, this results in the following cases

$$\frac{f(o \mid X, E_m)\mathbb{P}(E_m)}{f(o \mid X, E'_j)\mathbb{P}(E'_j)} \leq \begin{cases} \frac{2\omega}{W} e^{\frac{\epsilon_a}{2} + \epsilon_\mu + \frac{\epsilon}{2}}, & \text{if } o \text{ falls in an odd rank interval in} \mathcal{R}_X \text{ and an even rank interval} \\ e^{\frac{\omega}{b_{LEXP}} + \epsilon_\mu + \frac{\epsilon}{2}}, & \text{if } o \text{ falls in an odd rank interval in both cases,} \\ \frac{W}{2\omega} e^{\frac{\omega}{b_{LEXP}} - \frac{\epsilon_a}{2} + \frac{\epsilon}{2}}, & \text{if } o \text{ falls in an odd rank interval in } \mathcal{R}_X \text{ and an even rank interval} \\ e^{\frac{\epsilon_a}{2} + \frac{\epsilon}{2}}, & \text{if } o \text{ falls in an even rank interval in both } \mathcal{R}_X \text{ and } \mathcal{R}_{X'}. \end{cases}$$

The above holds since $|m - j| \leq 2$, which gives us:

$$\frac{f(o \mid X)}{f(o \mid X')} \leq \max\left(e^{\epsilon_a + \epsilon_\mu + \epsilon + \log(\frac{2\omega}{W})}, e^{\frac{\omega}{b_{LEXP}} + \epsilon_\mu + \frac{\epsilon_a}{2} + \epsilon}, e^{\frac{\omega}{b_{LEXP}} + \epsilon + \log(\frac{W}{2\omega})}, e^{\epsilon_a + \epsilon}\right)$$

Therefore, the mechanism Q_{GEXP} is

$$\max \begin{pmatrix} \epsilon_a + \epsilon_\mu + \epsilon + \log(\frac{2\omega}{W}) \\ \frac{\omega}{b_{LEXP}} + \epsilon_\mu + \frac{\epsilon_a}{2} + \epsilon \\ \frac{\omega}{b_{LEXP}} + \epsilon + \log(\frac{W}{2\omega}) \\ \epsilon_a + \epsilon \end{pmatrix}$$

-differentially private.

□

B Gaussian Randomization

B.1 Randomized data simple differentially private quantile mechanism - RQM

Definition B.1 (Laplace Exponential Mechanism)**.** Let $q \in (0,1)$, $a, b \in \mathbb{R}$, $\omega, b_{LEXP} > 0$, and $X^O \in [a,b]^n$ be a dataset.

We construct an extended dataset $X^o \in [a,b]^{n+2}$ by adding boundary values:

$$X^o_{(1)} = a, \quad X^o_{(n+2)} = b.$$

Next, we create a new vector $X \in \mathcal{R} = [a-(\lceil qn \rceil - 1)\omega, b+(n-\lceil qn \rceil - 1)\omega]^{n+2}$ by ensuring a distance of at least $2\omega > 0$ between samples.

Formally, for $k \geq 1$, we define:

$$X_{(r_q+k)} = X_{(r_q+k-1)} + t^+(k),$$
$$X_{(r_q-k)} = X_{(r_q-k+1)} + t^-(k).$$

where $t^+(k) = \max\left(2\omega, X^o_{(\lceil qn \rceil + k)} - X_{(r_q+k-1)}\right)$ and $t^-(k) = -\max\left(2\omega, X_{(r_q-k+1)} - X^o_{(\lceil qn \rceil - k)}\right)$

The dataset divides $\mathcal{R}$ into $2n+3$ intervals, denoted as I_k, defined as follows:

$$I_k = \begin{cases} [x_{(k)} - \omega, x_{(k)} + \omega], & \text{if } k \text{ is odd,} \\ [x_{(k)} + \omega, x_{(k)} - \omega], & \text{if } k \text{ is even.} \end{cases}$$

Let E_m denote the event of selecting an interval I_m from the range $\mathcal{R}_X$. Considers any points $\mu_k \in I_k$

The mechanism generates an output $o \in [a-(\lceil qn \rceil - 1)\omega, b+(n-\lceil qn \rceil - 1)\omega]$ with the following probability density:

$$f(o) = \sum_{I_k \in \mathcal{R}_X} f_{I_k}(o) \cdot \mathbb{P}(E_k),$$

where the density function $f_{I_k}(o)$ is determined by the interval I_k:

$$f_{I_k}(o) = \begin{cases} f(o \mid E_k) = f_{\mathcal{N}(\mu_k,\sigma^2)} \mathbb{1}_{I_k}(o), & \text{if } k \text{ is odd,} \\ \frac{1}{x_{(k+1)} - x_{(k)} - 2\omega} \mathbb{1}_{I_k}(o), & \text{if } k \text{ is even.} \end{cases}$$

The probability of selecting interval E_k is given by:

$$\mathbb{P}(E_k) \propto \begin{cases} (x_{(k+1)} - x_{(k)} - 2\omega) \cdot u(I_k), & \text{if } k \text{ is even,} \\ 2\omega \cdot u(I_k), & \text{if } k \text{ is odd.} \end{cases}$$

The utility function $u(I_k)$ is defined as:

$$u(I_k) = \begin{cases} \exp(\epsilon_\mu) \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is odd and } r(k) = 0, \\ \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is odd and } r(k) \neq 0, \\ \exp\left(-\frac{\epsilon_a}{2}\right) \exp\left(-\frac{\epsilon}{2} r(k)\right), & \text{if } k \text{ is even.} \end{cases}$$

Here, $r(k)$ is the rank function, defined as:

$$r(k) = \left| \left\lceil \frac{k}{2} \right\rceil - r_q \right|,$$

where:

- $u(I_k)$ is the utility function for interval I_k,
- q is the target quantile, and
- $r(k)$ is the rank of the sampled interval relative to the target quantile.

Theorem (Randomized Covariate Quantile Mechanism DP Guarantees). Let $\delta > 0$, choose ω s.t $\mathbb{P}_{\eta \sim \mathcal{N}(0,\sigma^2)}(|\eta| > \omega) \leq \delta$

Sampling a value with the density defined in 5.2 satisfies approximate

$$\max \begin{pmatrix} \epsilon_a + \epsilon_\mu + \epsilon + \log(\frac{\omega}{W}) \\ \frac{\omega^2}{\sigma^2} + \epsilon_\mu + \frac{\epsilon_a}{2} + \epsilon \\ \frac{\omega^2}{\sigma^2} + \epsilon + \log(\frac{W}{\omega}) \\ \epsilon + \epsilon_a \end{pmatrix}, \delta$$

-differential privacy.

where $W = \sqrt{2\pi}\sigma$,

B.1.1 Proof of theorem B.1

Proof. Let $X \sim X'$.

Consider an arbitrary output in the centered interval $[\hat{x}_{(k)} \pm t_k - 2\omega, \hat{x}_{(k)} \pm t_k + 2\omega]$, corresponding to a clipped sample drawn from $\mathcal{N}(\mu_k, \sigma^2)$, where the perturbed sample satisfies $\left|\mu_k - \hat{x}_{(k)}\right| \leq \omega$. This mechanism satisfies differential privacy, analogous to the Gaussian exponential mechanism, where the Gaussian distribution replaces its truncated counterpart.

Using the following inequality:

$$P\left(\left|\hat{x}_{(k)} - x_{(k)}\right| > d\right) \leq 2\sqrt{\frac{\sigma^2}{2\pi d^2}} e^{-\frac{d^2}{2\sigma^2}},$$

we select d such that $P\left(\left|\hat{x}_{(k)} - \hat{x}_{(k)}\right| \geq d\right) \leq \delta$. We then use the Gaussian exponential mechanism with $\omega \geq d$.

This ensures that, with probability $1 - \delta$, the means of the distributions $x_{(k)} \pm t_k$ and the independent sample $\hat{x}_{(k)} \pm t_k \sim \mathcal{N}(x_{(k)} \pm t_k, \sigma^2)$ lie in the same interval, i.e., $\left|x_{(k)} - \hat{x}_{(k)}\right| \leq \omega$.

If a centered interval is sampled and such d is found, the perturbed sample can be released while satisfying a level of differential privacy:

$$\epsilon = \max \begin{pmatrix} \epsilon_a + \epsilon_\mu + \epsilon + \log\left(\frac{\omega}{W}\right), \\ \frac{\omega^2}{\sigma^2} + \epsilon_\mu + \frac{\epsilon_a}{2} + \epsilon, \\ \frac{\omega^2}{\sigma^2} + \epsilon + \log\left(\frac{W}{\omega}\right), \\ \epsilon + \epsilon_a \end{pmatrix} \text{-differential privacy.}$$

The proof proceeds in a manner similar to A.2.1, differing only in the normalization ratio for the Gaussian mechanism and the bounds for the following:

$$p_1, p_2 \in [2n+3], \quad p_1 \equiv 1 \pmod 2, \quad p_2 \equiv 1 \pmod 2, \quad |p_1 - \lceil qn \rceil| = 0.$$

For such p_1 and p_2, we have:

$$\frac{f(o \mid E_{p_1})\mathbb{P}(E_{p_1})}{f(o \mid E_{p_2})\mathbb{P}(E_{p_2})} \leq e^{2\frac{\omega^2}{2\sigma^2} + \epsilon_\mu}.$$

□

C Consistency guarantees

C.1 proof of lemma 6.1

Lemma. For any $q \in (0, 1)$. The following consistency result holds for the sample quantile as defined in 2.1

$$\forall t > 0 \quad \mathbb{P}(|q_s - q_p| > t) \leq 2e^{-2nt^2 d_q^2}$$

Proof. Let $t > 0$. We start with:

$$\begin{aligned}
&\mathbb{P}(q_s - q_p > t) = \mathbb{P}\left(\sum_{i=1}^{n} \mathbb{1}_{x_i > q_p + t} \geq (1-q)n\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \left(\mathbb{1}_{x_i > q_p + t} - \mathbb{E}(\mathbb{1}_{x_i > q_p + t})\right) \geq (1-q)n - \sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_{x_i > q_p + t})\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \left(\mathbb{1}_{x_i > q_p + t} - \mathbb{E}(\mathbb{1}_{x_i > q_p + t})\right) \geq (1-q)n - \sum_{i=1}^{n} \mathbb{P}(x_i > q_p + t)\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \left(\mathbb{1}_{x_i > q_p + t} - \mathbb{E}(\mathbb{1}_{x_i > q_p + t})\right) \geq (1-q)n - \sum_{i=1}^{n} \left[1 - F\left(q_p + t\right)\right]\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \left(\mathbb{1}_{x_i > q_p + t} - \mathbb{E}(\mathbb{1}_{x_i > q_p + t})\right) \geq (1-q)n - n + nF\left(q_p + t\right)\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \left(\mathbb{1}_{x_i > q_p + t} - \mathbb{E}(\mathbb{1}_{x_i > q_p + t})\right) \geq nF\left(q_p + t\right) - qn\right)
\end{aligned}$$

The first equality comes from the fact that if the sample quantile is greater than the value $q_p + t$, then we at least sampled $(1-q)n$ points above that value.

We also have, by definition of d_q:

$$\begin{aligned}
F(q_p + t) &= F(q_p) + \int_{q_p}^{q_p + t} f(u)\, du \\
&\geq q + d_q t
\end{aligned}$$

The random variable $\mathbb{1}_{x_i > q_p + t}$ takes values in $[0, 1]$, so we can apply the Hoeffding inequality [Hoe94].

$$\begin{aligned}
\mathbb{P}(q_s > q_p + t) &\leq e^{-2n(F(q_p + t) - q)^2} \\
&\leq e^{-2nt^2 d_q^2}
\end{aligned}$$

Similarly, it also holds that if q_s is less than the value $q_p - t$, then we at least sampled qn elements below that value:

$$
\begin{aligned}
\mathbb{P}(q_s - q_p < -t) &= P\left(\sum_{i=1}^{n} \mathbb{1}_{x_i < q_p - t} \geq qn\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \left(\mathbb{1}_{x_i < q_p - t} - \mathbb{E}(\mathbb{1}_{x_i < q_p - t})\right) \geq qn - \sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_{x_i < q_p - t})\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \left(\mathbb{1}_{x_i < q_p - t} - \mathbb{E}(\mathbb{1}_{x_i < q_p - t})\right) \geq qn - \sum_{i=1}^{n} \mathbb{P}(x_i < q_p - t)\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{n} \left(\mathbb{1}_{x_i < q_p - t} - \mathbb{E}(\mathbb{1}_{x_i < q_p - t})\right) \geq qn - nF\left(q_p - t\right)\right)
\end{aligned}
$$

We observe that

$$
\begin{aligned}
F(q_p - t) = F(q_p) - \int_{q_p - t}^{q_p} f(u)\, du &\leq q - d_q t \\
qn - nF\left(q_p - t\right) &\geq n d_q t
\end{aligned}
$$

Therefore, we have

$$
\mathbb{P}(|q_s - q_p| > t) \leq 2e^{-2nt^2 d_q^2}.
$$

□

C.2 Proof of lemma 6.2

Lemma. Let $\omega > 0$ be defined as in 4. Fix $t > \omega$, and define $d_t \triangleq \inf_{u \in [q_s, q_s + t]} f(u)$.

$$
\begin{aligned}
\mathbb{P}\left(Q_{\mathrm{CEXP}} - q_s > t\right) &\leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) e^{-\frac{(n(1-q)td_t)\epsilon}{2}} + e^{-\frac{td_t q^2}{2} n} \\
\mathbb{P}\left(Q_{\mathrm{GEXP}} - q_s > t\right) &\leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) e^{-\frac{(n(1-q)td_t)\epsilon}{2}} + e^{-\frac{td_t q^2}{2} n}
\end{aligned}
$$

where $C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) \triangleq \frac{1}{4\omega} e^{\frac{\epsilon - 2\epsilon_\mu}{4}} \left(1 + e^{-\epsilon_a/2}\right)$.

Proof. Fix $q \in (0, 1)$. recall that $d_t \triangleq \inf_{u \in [q_s, q_s + t]} f$, which is a well defined quantity if we slightly perturb the samples with Gaussian noise. Let E be the event that there exist at least $n_t = n(1-q)td_t \leq (1-q)n$ points between the sample quantile q_s and $q_s + t$. It is clear that this number can be at most $(1-q)n$. the term td_t can be seen as a lower bound on the proportion of element in the interval $[q_s, q_s + t]$ from a population standpoint, therefore by multiplying by the total number to the right of the target quantile, the constant n_t can be seen as a lower bound on the number of elements in $[q_s, q_s + t]$.

Using the law of total probability, we have:

$$\begin{aligned}\mathbb{P}\left(Q_{\text{priv}} - q_s > t\right) &= \mathbb{P}\left(Q_{\text{priv}} - q_s > t \mid E\right)\mathbb{P}\left(E\right) \\ &\quad + \mathbb{P}\left(Q_{\text{priv}} - q_s > t \mid \overline{E}\right)\mathbb{P}\left(\overline{E}\right) \\ &\leq \mathbb{P}\left(Q_{\text{priv}} - q_s > t \mid E\right) + \mathbb{P}\left(\overline{E}\right).\end{aligned}$$

To upper bound $\mathbb{P}\left(\overline{E}\right)$, we first observe the following:

$$\int_{q_s}^{q_s+t} f(u)\,du \geq td_t.$$

Now we can write:

$$\begin{aligned}\mathbb{P}\left(\overline{E}\right) &= P\left(\sum_{i\in[n]} \mathbb{1}_{q_s \leq x_i \leq q_s+t} < n_t\right) \\ &= \mathbb{P}\left(\text{Binom}\left(n, \int_{q_s}^{q_s+t} f(u)\,du\right) < n(1-q)td_t\right) \\ &\leq e^{-\frac{td_t q^2}{2}n}\end{aligned}$$

Above, we used independence of samples and the Chernoff bound for the binomial distribution which we recall in the appendix F.

When we condition on the event E, we have that for any $s \geq q_s + t$, the rank of the interval it will fall in satisfies

$$\left|\lceil\frac{i}{2}\rceil - qn\right| \geq n_t.$$

In other words, there are at least $2 \times n_t$ intervals before the point $s \in [q_s + t, b]$.

We start by lower bounding the normalization factor by only considering the odd rank intervals. We will consider $x_{(1)} = (a - \lceil qn\rceil - 1)\,\omega$ and $x_{(n)} = b + (n - \lceil qn\rceil)\omega$. It holds that

$$\begin{aligned}\int_{x_{(1)}}^{x_{(n)}} u_{CEXP}\,dx &\geq 2\omega \sum_{k=1}^{n-\lceil qn\rceil} e^{-\frac{\epsilon}{2}k} + 2\omega e^{\epsilon_\mu} \\ &\geq 2\omega(e^{-\frac{\epsilon}{2}} + e^{\epsilon_\mu}) \\ &\geq 4\omega e^{\frac{2\epsilon_\mu - \epsilon}{4}}.\end{aligned}$$

We further consider events Od denoting sampling from an odd rank interval and the event Ev that is realized when we select an even rank interval, using $P\left(Od\right), P\left(Ev\right) \leq 1$:

$$\mathbb{P}\left(Q_{\text{priv}} - q_s > t\right) \leq \mathbb{P}\left(Q_{\text{priv}} - q_s > t \mid E, Od\right) + \mathbb{P}\left(Q_{\text{priv}} - q_s > t \mid E, Ev\right) + \mathbb{P}\left(\overline{E}\right).$$

We also use the lower bound on the interval rank to write the following:

$$\mathbb{P}\left(Q_{\text{priv}} - q_s > t, E, Ev\right) < \frac{e^{-\frac{\epsilon_a}{2}} e^{-\frac{n(1-q)td_t\epsilon}{2}}}{4\omega e^{\frac{2\epsilon_\mu - \epsilon}{4}}},$$
$$\mathbb{P}\left(Q_{\text{priv}} - q_s > t, E, Od\right) < \frac{e^{-\frac{n(1-q)td_t\epsilon}{2}}}{4\omega e^{\frac{2\epsilon_\mu - \epsilon}{4}}}.$$

Above we used $t > \omega$, This gives us the final bound:

$$\mathbb{P}\left(Q_{\text{priv}} - q_s > t\right) \leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) e^{-\frac{n(1-q)td_t\epsilon}{2}} + e^{-\frac{td_tq^2}{2}n}.$$

s.t. $C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) \triangleq \frac{1}{4\omega} e^{-\frac{\epsilon - 2\epsilon_\mu}{4}} \left(1 + e^{-\epsilon_a/2}\right)$.

□

C.3 Proof of lemma 6.3

Lemma. Let $\omega > 0$ be defined as in 4. Fix $t > \omega$, $q_s = X_{\lceil qn \rceil}$ such that $\exists \eta > 0$

$$\mathbb{E}\left(\mathbb{1}_{x \sim P \in [q_s, q_s + t]}\right) \geq \eta n.$$

Then

$$P\left(Q_{\text{CEXP}} - q_s > t\right) \leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) \cdot e^{-\frac{\left(\mathbb{E}\left(1_{x \sim P \in [q_s, q_s + t]}\right) - \eta n\right)\epsilon}{2}} + e^{-2n\eta^2}$$
$$P\left(Q_{\text{GEXP}} - q_s > t\right) \leq C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right) \cdot e^{-\frac{\left(\mathbb{E}\left(1_{x \sim P \in [q_s, q_s + t]}\right) - \eta n\right)\epsilon}{2}} + e^{-2n\eta^2}$$

where $C\left(\epsilon, \epsilon_\mu, \epsilon_a, \omega\right)$ is defined as in 6.2.

The term η is a tolerance parameter. The private estimator will produce fewer outputs in the range $[q_s, q_s + t]$ as $\mathbb{E}\left(\mathbb{1}_{x \sim P \in [q_s, q_s + t]}\right)$ gets larger.

Proof. note the existence of the constant η is guaranteed by addition of gaussian noise without any prior assumptions on the distribution, we proceed similar to the previous lemma by considering

$E = \{\sum_{i \in [n]} \mathbb{1}_{x \sim P \in [q_s, q_s + t]} \geq \sum_{i \in [n]} \mathbb{E}\left(\mathbb{1}_{x \sim P \in [q_s, q_s + t]}\right) - \eta n\}$

Conditioning on this event allows to provide a lower bound on the rank i of elements in $[q_s+t, b+(n-\lceil qn \rceil)]$ in the output space of the centered exponential mechanism.

we can say that $|i-qn| \geq \left(\sum_{i\in[n]} \mathbb{E}\left(\mathbb{1}_{x\sim P\in[q_s,q_s+t]}\right) - \eta n\right)$

this allows us to find the leading term similar to 6.3 We conclude by using A hoeffding inequality on the random variable to compute $P\left(\overline{E}\right)$

□

D Unbiased differentially private quantile release mechanism - UBEXP

To simplify notation we will use r_q and $\lceil qn \rceil$ interchangeably to refer to the rank of the target quantile under datasets $X^O, andX$

D.1 proof of lemma 7.2

Lemma. Let $X^O, X^{O'} \in [a,b]^n$ be two adjacent datasets, with $X^{O'}$ created by changing a value to a value larger larger than it. Let $X^o, X^{o'} \in [a,b]^{2mn_q+3}$, $X, X' \in \mathcal{R}^{2mn_q+3}$ be vectors constructed by the translations defined in Definition 7.1, $r_q = mn_q + 2$ the quantile rank under X. Define mappings $s_{l,X}, s_{r,X} : [1, mn_q] \to \mathbb{R}$ as follows:

$$
\begin{aligned}
s_{l,X}(k) &= x^o_{(\lceil qn \rceil)} - x^o_{(\lceil qn \rceil - k)}, \\
s_{r,X}(k) &= x^o_{(\lceil qn \rceil + k)} - x^o_{(\lceil qn \rceil)}, \\
s_{l,X'}(k) &= x^{o'}_{(\lceil qn \rceil)} - x^{o'}_{(\lceil qn \rceil - k)}, \\
s_{r,X'}(k) &= x^{o'}_{(\lceil qn \rceil + k)} - x^{o'}_{(\lceil qn \rceil)}.
\end{aligned}
$$

Both mappings $s_{l,X}$ and $s_{r,X}$ are non-decreasing, with the following properties holding for all $k \in [0, mn_q]$:

$$
x^o_{(\lceil qn \rceil + k)} \leq x^{o'}_{(\lceil qn \rceil + k)} \leq x^o_{(\lceil qn \rceil + k + 1)},
$$

and for all $k \in [1, mn_q]$:

$$
x^o_{(\lceil qn \rceil - k)} \leq x^{o'}_{(\lceil qn \rceil - k)} \leq x^o_{(\lceil qn \rceil - k + 1)}.
$$

Furthermore, for any $m \geq 1$ and $j \geq 0$, the following inequalities hold:

$$
x_{(r_q+j)} \geq x^{'}_{(r_q-m)}, \quad x^{'}_{(r_q+j)} \geq x_{(r_q-m)}.
$$

Proof. We begin by establishing the properties of the mappings defined on the interval $\{1, \ldots, mn_q + 1\}$. These mappings are non-decreasing due to the definition of the translations, taking larger steps with each increment.

To create the adjacent dataset X', for any $k \in [n]$, if we change a value to a larger value as established in 4.2, the resulting dataset will satisfy:

$$x^{o,'}_{(k)} \geq x^{o}_{(k)}.$$

Since we always include the endpoints a and b to construct $X^o \in [a,b]^{2mn_q+3}$, the modified sequence maintains the ordering for any $k \in [0, mn_q]$:

$$x^{o,'}_{(k)} \leq x^{o}_{(k+1)}.$$

Thus, any change remains bounded within the maximum value by construction.

This leads us to conclude:

$$\forall k \in [|0, mn_q|], \quad x^{o}_{(qn+k)} \leq x^{o,'}_{(qn+k)} \leq x^{o}_{(qn+k+1)},$$

$$\forall k \in [|1, mn_q+1|], \quad x^{o}_{(qn-k)} \leq x^{o,'}_{(qn-k)} \leq x^{o}_{(qn-k+1)}.$$

Let $\text{Gap}_q = d = x^{o,'}_{\lceil qn \rceil} - x^{o}_{\lceil qn \rceil} \leq b - a$

Now, to prove the remaining properties, let $m \geq 1, j \geq 0$. We have:

$$\begin{aligned}
&x'_{(\lceil qn \rceil - m)} \leq x'_{(\lceil qn \rceil - 1)} = x^{o,'}_{(\lceil qn \rceil)} - \text{Gap}_q - \max\left(x^{o,'}_{\lceil qn \rceil} - x^{o,'}_{\lceil qn \rceil - 1},\, x^{o,'}_{\lceil qn \rceil + 1} - x^{o,'}_{\lceil qn \rceil}\right), \\
&x'_{(\lceil qn \rceil - m)} \leq x^{o,'}_{(\lceil qn \rceil)} - \text{Gap}_q, \\
&x'_{(\lceil qn \rceil - m)} \leq x^{o}_{(\lceil qn \rceil)}, \\
&x'_{(\lceil qn \rceil - m)} \leq x_{(\lceil qn \rceil + j)}.
\end{aligned}$$

Above, we used the property from the previous lemma $x^{o}_{\lceil qn \rceil} \geq x^{o,'}_{\lceil qn \rceil - 1}$. Additionally, we have:

$$\begin{aligned}
&x_{(\lceil qn \rceil - m)} \leq x'_{(\lceil qn \rceil - 1)} = x^{o}_{(\lceil qn \rceil)} - \text{Gap}_q - \max\left(x^{o}_{\lceil qn \rceil} - x^{o}_{\lceil qn \rceil - 1},\, x^{o}_{\lceil qn \rceil + 1} - x^{o}_{\lceil qn \rceil}\right), \\
&x_{(\lceil qn \rceil - m)} \leq x^{o}_{(\lceil qn \rceil)}, \\
&x_{(\lceil qn \rceil - m)} \leq x^{o,'}_{(\lceil qn \rceil)}, \\
&x_{(\lceil qn \rceil - m)} \leq x^{'}_{(\lceil qn \rceil + j)}.
\end{aligned}$$

$$\begin{aligned}
&\forall m \geq 1, j \geq 0, \quad x'_{(\lceil qn \rceil - m)} \leq x^{o}_{(\lceil qn \rceil + 1)} \leq x_{(\lceil qn \rceil + j)}, \\
&\forall m \geq 1, j \geq 0, \quad x_{(\lceil qn \rceil - m)} \leq x^{o,'}_{(\lceil qn \rceil)} \leq x'_{(\lceil qn \rceil + j)}.
\end{aligned}$$

□

D.2 proof of lemma 7.3

Lemma. Let $X^O, X^{O'} \in [a,b]^n$ be two adjacent datasets, with $X^{O'}$ created by changing a value to a larger value. Let $X^o, X^{o'} \in [a,b]^{2mn_q+3}, X, X' \in \mathcal{R}^{2mn_q+3}$ be vectors constructed by the translations defined in Definition 7.1, $r_q = mn_q + 2$ the quantile rank under X.

For all $k \in [0, mn_q + 1]$, we have:

$$x_{(r_q+k)} \geq x'_{(r_q+k-2)},$$
$$x'_{(r_q+k)} \geq x_{(r_q+k-2)}.$$

Similarly, for all $k \in [2, mn_q + 1]$, we have:

$$x_{(r_q-k+2)} \geq x'_{(r_q-k)},$$
$$x'_{(r_q-k+2)} \geq x_{(r_q-k)}.$$

Proof. Let $X^O, X^{O,'} \in [a,b]^n$ be adjacent datasets, and let $X^o, X^{o,'} \in [a,b]^{2mn_q+3}, X, X'$ denote vectors constructed following Definition 7.1.

By Lemma 7.3, for $m = 1$ and $j = 1$, we have:

$$x_{(\lceil qn \rceil+1)} \geq x'_{(\lceil qn \rceil-1)},$$

and we also have,

$$x_{(\lceil qn \rceil+2)} \geq x^o_{(\lceil qn \rceil)} + \mathrm{Gap}_q \geq x^{o,'}_{(\lceil qn \rceil)} = x'_{(\lceil qn \rceil)}.$$

This confirms the inequality for any $k' \in \{0, 1\}$.

Next, for $k' \in [2, mn_q - 1]$, let $k = k' - 2 \geq 0$. We proceed with the following derivation:

$$\begin{aligned}
x_{(\lceil qn \rceil+k+2)} &= x^o_{(\lceil qn \rceil)} + (k+2)\mathrm{Gap}_q + \max\left(x^o_{(\lceil qn \rceil+k+2)} - x^o_{(\lceil qn \rceil)}, x^o_{(\lceil qn \rceil)} - x^o_{(\lceil qn \rceil-k-2)}\right) \\
&\geq x^o_{(\lceil qn \rceil)} + (k+2)\mathrm{Gap}_q + \max\left(x^{o,'}_{(\lceil qn \rceil+k+1)} - x^{o,'}_{(\lceil qn \rceil)}, x^o_{(\lceil qn \rceil)} - x^{o,'}_{(\lceil qn \rceil-k-2)}\right) \\
&\geq x^o_{(\lceil qn \rceil)} + (k+2)\mathrm{Gap}_q + \max\left(x^{o,'}_{(\lceil qn \rceil+k+1)} - x^{o,'}_{(\lceil qn \rceil)}, x^o_{(\lceil qn \rceil)} - x^{o,'}_{(\lceil qn \rceil-k-1)}\right) \\
&\geq x^o_{(\lceil qn \rceil)} + (k+2)\mathrm{Gap}_q + \max\left(x^{o,'}_{(\lceil qn \rceil+k+1)} - x^{o,'}_{(\lceil qn \rceil)}, x^o_{(\lceil qn \rceil)} - x^{o,'}_{(\lceil qn \rceil-k-1)}\right) \\
&\geq x^o_{(\lceil qn \rceil)} + (k+2)\mathrm{Gap}_q + \max\left(s_{r,X'}(k+1), s_{l,X'}(k+1) - \mathrm{Gap}_q\right) \\
&\geq x^o_{(\lceil qn \rceil)} + (k+2)\mathrm{Gap}_q - \mathrm{Gap}_q + \max\left(s_{r,X'}(k+1), s_{l,X'}(k+1)\right) \\
&\geq x^o_{(\lceil qn \rceil)} + (k+1)\mathrm{Gap}_q + \max\left(s_{r,X'}(k+1), s_{l,X'}(k+1)\right) \\
&\geq x^{o,'}_{(\lceil qn \rceil)} + k\mathrm{Gap}_q + \max\left(s_{r,X'}(k), s_{l,X'}(k)\right) \\
&\geq x'_{(\lceil qn \rceil+k)}.
\end{aligned}$$

Above we repeatedly used lemma 7.2.

Similarly,

$$
\begin{aligned}
x'_{(\lceil qn\rceil+k+2)} &= x^{o,'}_{(\lceil qn\rceil)} + (k+2)\mathrm{Gap}_q + \max\left(x^{o,'}_{(\lceil qn\rceil+k+2)} - x^{o,'}_{(\lceil qn\rceil)}, x^{o,'}_{(\lceil qn\rceil)} - x^{o,'}_{(\lceil qn\rceil-k-2)}\right)\\
x'_{(\lceil qn\rceil+k+2)} &\geq x^{o}_{(\lceil qn\rceil)} + (k+2)\mathrm{Gap}_q + \max\left(x^{o}_{(\lceil qn\rceil+k+1)} - x^{o,'}_{(\lceil qn\rceil)}, x^{o}_{(\lceil qn\rceil)} - x^{o}_{(\lceil qn\rceil-k-1)}\right)\\
x'_{(\lceil qn\rceil+k+2)} &\geq x^{o}_{(\lceil qn\rceil)} + (k+2)\mathrm{Gap}_q + \max\left(s_{r,X}\left(k+1\right) - \mathrm{Gap}_q, +s_{l,X}\left(k+1\right)\right)\\
x'_{(\lceil qn\rceil+k+2)} &\geq x^{o}_{(\lceil qn\rceil)} + (k+1)\mathrm{Gap}_q + \max\left(s_{r,X}\left(k+1\right), s_{l,X}\left(k+1\right)\right)\\
x'_{(\lceil qn\rceil+k+2)} &\geq x^{o}_{(\lceil qn\rceil)} + k\mathrm{Gap}_q + \max\left(s_{r,X}\left(k\right), s_{l,X}\left(k\right)\right)\\
x'_{(\lceil qn\rceil+k+2)} &\geq x_{(\lceil qn\rceil+k)}
\end{aligned}
$$

Hence, for all $k \in [0, mn_q - 1]$, we conclude:

$$
\begin{aligned}
x_{(\lceil qn\rceil+k+2)} &\geq x'_{(\lceil qn\rceil+k)},\\
x'_{(\lceil qn\rceil+k+2)} &\geq x_{(\lceil qn\rceil+k)}.
\end{aligned}
$$

Now we want to prove that for $k \in [2, mn_q + 1]$, we have:

$$
\begin{aligned}
x_{(\lceil qn\rceil-k+2)} &\geq x'_{(\lceil qn\rceil-k)},\\
x'_{(\lceil qn\rceil-k+2)} &\geq x_{(\lceil qn\rceil-k)}.
\end{aligned}
$$

$$
\begin{aligned}
x_{(\lceil qn\rceil-k+2)} &= x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(x^{o}_{(\lceil qn\rceil+k-2)} - x^{o}_{(\lceil qn\rceil)},\ x^{o}_{(\lceil qn\rceil)} - x^{o}_{(\lceil qn\rceil-k+2)}\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(x^{o,'}_{(\lceil qn\rceil+k-2)} - x^{o}_{(\lceil qn\rceil)},\ x^{o,'}_{(\lceil qn\rceil)} - x^{o,'}_{(\lceil qn\rceil-k)}\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(x^{o,'}_{(\lceil qn\rceil+k-2)} - x^{o}_{(\lceil qn\rceil)},\ s_{l,X'}(k)\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(x^{o,'}_{(\lceil qn\rceil+k)} - x^{o}_{(\lceil qn\rceil)},\ s_{l,X'}(k)\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(s_{r,X'}(k) + \mathrm{Gap}_q,\ s_{l,X'}(k)\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \mathrm{Gap}_q + \max\left(s_{r,X'}(k),\ s_{l,X'}(k)\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-1)\mathrm{Gap}_q - \max\left(s_{r,X'}(k),\ s_{l,X'}(k)\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x^{o,'}_{(\lceil qn\rceil)} - \mathrm{Gap}_q - (k-1)\mathrm{Gap}_q - \max\left(s_{r,X'}(k),\ s_{l,X'}(k)\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x^{o,'}_{(\lceil qn\rceil)} - k\mathrm{Gap}_q - \max\left(s_{r,X'}(k),\ s_{l,X'}(k)\right)\\
x_{(\lceil qn\rceil-k+2)} &\geq x_{(\lceil qn\rceil-k)}
\end{aligned}
$$

We used $x^{o}_{(\lceil qn\rceil-k+2)} \geq x^{o}_{(\lceil qn\rceil-k+1)} \geq x^{o}_{(\lceil qn\rceil-k)}$.
Similarly, we have

$$
\begin{aligned}
x'_{(\lceil qn\rceil-k+2)} &= x^{o,'}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(x^{o,'}_{(\lceil qn\rceil+k-2)} - x^{o,'}_{(\lceil qn\rceil)},\ x^{o,'}_{(\lceil qn\rceil)} - x^{o,'}_{(\lceil qn\rceil-k+2)}\right)\\
x'_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(x^{o}_{(\lceil qn\rceil+k-1)} - x^{o}_{(\lceil qn\rceil)},\ x^{o,'}_{(\lceil qn\rceil)} - x^{o}_{(\lceil qn\rceil-k+2)}\right)\\
x'_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(x^{o}_{(\lceil qn\rceil+k-1)} - x^{o}_{(\lceil qn\rceil)},\ x^{o,'}_{(\lceil qn\rceil)} - x^{o}_{(\lceil qn\rceil-k+1)}\right)\\
x'_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \max\left(s_{r,X}(k-1),\ s_{l,X}(k-1) + \mathrm{Gap}_q\right)\\
x'_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-2)\mathrm{Gap}_q - \mathrm{Gap}_q + \max\left(s_{r,X}(k-1),\ s_{l,X}(k-1)\right)\\
x'_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-1)\mathrm{Gap}_q - \max\left(s_{r,X}(k-1),\ s_{l,X}(k-1)\right)\\
x'_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - (k-1)\mathrm{Gap}_q - \max\left(s_{r,X}(k),\ s_{l,X}(k)\right)\\
x'_{(\lceil qn\rceil-k+2)} &\geq x^{o}_{(\lceil qn\rceil)} - k\mathrm{Gap}_q - \max\left(s_{r,X}(k),\ s_{l,X}(k)\right)\\
x'_{(\lceil qn\rceil-k+2)} &\geq x_{(\lceil qn\rceil-k)}
\end{aligned}
$$

□

D.3 proof of Theorem D.1

Theorem D.1 (UBCEXP DP Guarantees)**.** *7.1 satisfies pure ϵ differential privacy*

Proof. Let $o \in \mathcal{R}$. In the lemma 7.3, we showed that the rank of the interval containing o in $\mathcal{R}_X$ and $\mathcal{R}_{X'}$ can change by at most 2.

We consider chaging a smaller value in X^O to a larger value to make the adjacent dataset, this comes at no loss of generality since we can always exchange datasets

we consider $o \in I_k = [x_{(k)}, x_{(k+1)})$

Note that the maximum and the minimum of the range are independent of the dataset abd by convention we can set any value outside the boundaties to the maximum or the minimum, by lemma 7.3 we can write

$$x_{(\lceil qn\rceil+k-2)} \geq x_{(\lceil qn\rceil+k)} \leq o \leq x_{(\lceil qn\rceil+k+1)} \leq x_{(\lceil qn\rceil+k+3)}$$

this means that this that the output $o \in \{I'_{k-2}, I'_k, I'_{k+2}\}$

this means that the rank changed by at most 2

The intervals created by $\mathcal{R}_X$ are distinct, Let j and j' denote the indices of the intervals containing o in $\mathcal{R}_X$ and $\mathcal{R}_{X'}$, respectively, we established that $|j' - j| = 2$. Then, we have:

$$
\begin{aligned}
\frac{f(o \mid X)}{f(o \mid X')} &\leq \frac{u(I_j)}{u(I'_j)} R_{x'|x}\\
\frac{f(o \mid X)}{f(o \mid X')} &\leq e^{\frac{\epsilon}{2}} R_{x'|x}
\end{aligned}
$$

$$R_{x'|x} = \frac{\sum_{I_k \in \mathcal{R}_X} \int u(o, X')\, do}{\sum_{I_k \in \mathcal{R}_X} \int u(o, X)\, do}$$
$$R_{x'|x} \leq e^{\frac{\epsilon}{2}} \frac{\sum_{I_k \in \mathcal{R}_X} \int u(o, X)\, do}{\sum_{I_k \in \mathcal{R}_X} \int u(o, X)\, do}$$
$$R_{x'|x} \leq e^{\frac{\epsilon}{2}}.$$

Thus, we conclude that for any output in the range of the mechanism UBCEXP the ratio of the densities is bounded by e^{ϵ}.

□

D.4 proof of Theorem 7.5

Theorem (Bias of UBCEXP). Let $X^O \in [a, b]^n$, $r_q = mn_q + 2$ the quantile rank under X. The mechanism 7.1 with the adjusted utility 1 is unbiased for differentially private median release under the swap adjacency relation. $\mathbb{E}\left(Q_{CEXP}\left(X^O\right)\right) = x_{r_q} = x^O_{(\lceil qn \rceil)}$

Proof. Let $X^o \in [a, b]^n$, and let X be the vector created by UBCEXP (Definition 7.1).

We construct $2mn_q + 3$ points and we have $2mn_q + 2$ intervals. We will denote the probability mass of an interval by $k \geq 1$ for $k \in [1, mn_q + 1]$, such that:

$$p\left(E_{(\lceil qn \rceil \pm k)}\right)$$

it holds that $p\left(E_{(\lceil qn \rceil + k)}\right) = p\left(E_{(\lceil qn \rceil - k)}\right)$

By the construction detailed in UBCEXP, we have:

$$\forall k \in [0, mn_q + 1], \quad x_{(\lceil qn \rceil + k)} + x_{(\lceil qn \rceil - k)} = 2x^O_{(\lceil qn \rceil)} = x_{(r_q)}$$

Now, for the expected value of Q_{UBCEXP}, we have:

$$\begin{aligned}
\mathbb{E}(Q_{UBCEXP}) = &\sum_{k=1}^{mn_q+1} \mathbb{P}(E_{\lceil qn \rceil + k}) \frac{x_{(\lceil qn \rceil + k)} + x_{(\lceil qn \rceil + (k-1))}}{2} \\
&+ \sum_{k=1}^{mn_q+1} \mathbb{P}(E_{\lceil qn \rceil - k}) \frac{x_{(\lceil qn \rceil - k)} + x_{(\lceil qn \rceil - (k-1))}}{2} \\
= &\sum_{k=1}^{mn_q+1} \mathbb{P}(E_{\lceil qn \rceil + k}) \frac{x_{(\lceil qn \rceil + k)} + x_{(\lceil qn \rceil - k)}}{2} \\
&+ \sum_{k=1}^{mn_q+1} \mathbb{P}(E_{\lceil qn \rceil - k}) \frac{x_{(\lceil qn \rceil + k - 1)} + x_{(\lceil qn \rceil - (k-1))}}{2} \\
= &\, x^O_{(\lceil qn \rceil)} = x_{(r_q)}.
\end{aligned}$$

□

E Experiements

For integer $j > 1$, let $Q = (q_1, q_2, \ldots, q_j) \in (0,1)^j$ and let $\mathcal{R}$ be the range of the mechanism and $o \in \mathcal{R}^j$ be the vector containing the dp estimators for each quantile, be the quantiles we want to release. We will first evaluate the proposed mechanism's performance in the case where no measurement noise is assumed on the dataset.

As the theoretical consistency guarantees are provided for the absolute value of the difference between the quantile estimator and the population / quantile it is more suitable to use it as a performance metric.

$$\mathrm{Perf}_m x\left(j, Q, o\right) = \max_{k \in [j]} \left| o_{(k)} - x_{(\lceil q_k n \rceil)} \right|.$$

Remark E.1. An alternative metric frequently used to assess quantile release algorithms is the count difference [GJK21, KSS22], which measures the proportion of datapoints on each side of the estimated quantile. However, we find that this metric does not adequately capture the theoretical issue we raised: if there is a significant gap between the sample quantile and nearby datapoints, an estimate that is far from the sample quantile could still separate datapoints similarly.

E.1 Note on numerical stability

For large datasets, the probability assigned to intervals will exponentially decay, yielding multiple entries with probability zero due to numerical limitations. To provide realistic results, it is more convenient to use mathematical properties of the exponential distribution that yield numerically stable results.

This issue has also been noted in papers studying differentially private quantile release with the exponential mechanism [DGHM$^+$22, MG20]. Both works relied on adding Gumbel-distributed noise to the interval probabilities before sampling, selecting the maximum value.

This is the approach we use in the experiments presented in the main paper.

E.2 Unbiased exponential mechanism - experiments

In this section, we evaluate the unbiased exponential mechanism, which is governed by the upper bound Gap_q. This mechanism makes no assumptions about the underlying distribution, but achieving zero bias requires expanding the clipping endpoints by $\mathcal{O}(n^2(b-a))$. We explore whether the unbiased exponential mechanism outperforms alternative methods for specific data distributions, particularly when Gap_q can be tightly bounded with high probability. To do so, we utilize Lemma F, formally stated and proven in the supplementary materials in the appendix, and demonstrate the mechanism with data generated from a uniform distribution.

To keep the probability $\delta = \mathcal{O}\left(\frac{1}{n}\right)$ small, the unbiased mechanism achieves comparable performance to the exponential mechanism but performs less ef-

fectively than the centered exponential mechanism, which focuses on reducing conditional bias.

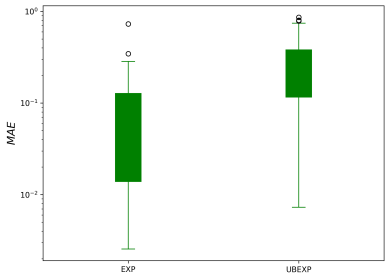

Figure 4: Results for $n = 10^3$, $\epsilon = 4$, $\delta = 10^{-3}$, $X \sim \text{Unif}(-10^3; 10^3)$.

F Technical complement

Definition F.1 (Laplace Mechanism, [DR+14])**.** The Laplace mechanism is a mechanism that provides ϵ-differential privacy by adding noise from a Laplace distribution to the output of a function. Given a function $f : \mathcal{D} \to \mathbb{R}^d$, the Laplace mechanism outputs a noisy version of $f(D)$

$$M(D) = f(D) + \text{Lap}\left(\frac{\Delta f}{\epsilon}\right),$$

where Δ is defined as in 2.3 and $\text{Lap}(\lambda)$ denotes the Laplace distribution with mean 0 and scale λ, i.e., the probability density function of $\text{Lap}(\lambda)$ is:

$$\mathbb{P}(x) = \frac{1}{2\lambda} \exp\left(-\frac{|x|}{\lambda}\right).$$

Definition F.2 (Gaussian Mechanism, [DR+14])**.** The Gaussian mechanism is a mechanism that provides (ϵ, δ)-differential privacy by adding noise from a Gaussian distribution to the output of a function. Given a function $f : \mathcal{D} \to \mathbb{R}^d$, the Gaussian mechanism outputs a noisy version of $f(D)$

$$M(D) = f(D) + \mathcal{N}\left(0, \sigma^2 I_d\right),$$

For (ϵ, δ)-differential privacy, the noise is drawn from a Gaussian distribution with a standard deviation σ that is calibrated to satisfy the privacy guarantee:

$$\sigma = \frac{\Delta_2 f}{\epsilon} \cdot \sqrt{2 \ln\left(\frac{1.25}{\delta}\right)}.$$

where Δ_2 is defined as in 2.3

Lemma (Chernoff Bound for a Binomial Distribution). Let $X \sim \text{Bin}(n, p)$ and let $\mu = \mathbb{E}[X]$. For any $0 < \delta < 1$, the Chernoff bound provides the following bounds:

- **Upper tail:** The probability that X exceeds $(1+\delta)\mu$ is bounded by

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \exp\left(-\frac{\delta^2\mu}{3}\right).$$

- **Lower tail:** The probability that X falls below $(1-\delta)\mu$ is bounded by

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq \exp\left(-\frac{\delta^2\mu}{2}\right).$$

where $\exp(x) = e^x$.

Let $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$

Lemma (Bernstein inequality). For any random variable X bounded by $|X - \mathbb{E}[X]| \leq b$, the following inequality holds for any $d > 0$:

$$\mathbb{P}((|X - \mathbb{E}[X]| \geq d) \leq 2\exp\left(-\frac{d^2}{2\,\text{Var}(X) + \frac{2bd}{3}}\right).$$

Lemma (Hoeffding's inequality [Hoe94]). For a bounded random variable $a \leq X \leq b$, the following inequality holds for any $d > 0$:

$$\mathbb{P}((|X - \mathbb{E}[X]| \geq d) \leq 2\exp\left(-2d^2\right).$$

Lemma (Gap Bound - uniform distribution). Let $X_1, \ldots, X_n \sim \text{Unif}(a, b)$ be random variables, and consider an integer $m \in [n-1]$.

With probability $1 - \delta$, the following holds:

$$|X_{(m+1)} - X_{(m)}| < 2d(b-a) + \frac{b-a}{n+1}$$

where $\delta = \min\left(2\exp\left(-\frac{d^2}{2\frac{(i+1)(n-i)}{(n+1)^2(n+2)} + \frac{2d}{3}}\right) + 2\exp\left(-\frac{d^2}{2\frac{i(n-i+1)}{(n+1)^2(n+2)} + \frac{2d}{3}}\right), 2\exp\left(-2d^2\right)\right)$

where $\text{Var}(X)$ is the variance of X.

Proof. Let $i \in [n-1]$, and define

$$U_i := \frac{X_i - a}{b-a} \sim \text{Unif}(0, 1),$$

where U_i is the normalized version of the random variable X_i. The order statistic $U_{(i)}$ follows the distribution $U_{(i)} \sim \text{Beta}(i, n-i+1)$.

Using Bernstein's inequality, we can obtain bounds for $U_{(i+1)}$ and $U_{(i)}$. Specifically,

$$\mathbb{P}(\left(\left|U_{(i+1)} - \frac{i+1}{n+1}\right| \geq d\right) \leq 2\exp\left(-\frac{d^2}{2\frac{(i+1)(n-i)}{(n+1)^2(n+2)} + \frac{2d}{3}}\right),$$

and

$$\mathbb{P}(\left(\left|U_{(i)} - \frac{i}{n+1}\right| \geq d\right) \leq 2\exp\left(-\frac{d^2}{2\frac{i(n-i+1)}{(n+1)^2(n+2)} + \frac{2d}{3}}\right).$$

Using the union bound, we get the following probability bound:

$$\mathbb{P}(\left(\overline{\left|U_{(i)} - \frac{i}{n+1}\right| < d \text{ and } \left|U_{(i+1)} - \frac{i+1}{n+1}\right| < d}\right) \leq \mathbb{P}(\left(\left|U_{(i+1)} - \frac{i+1}{n+1}\right| \geq d\right) + \mathbb{P}(\left(\left|U_{(i)} - \frac{i}{n+1}\right| \geq d\right).$$

Thus, with probability greater than

$$1 - 2\exp\left(-\frac{d^2}{2\frac{(i+1)(n-i)}{(n+1)^2(n+2)} + \frac{2d}{3}}\right) - 2\exp\left(-\frac{d^2}{2\frac{i(n-i+1)}{(n+1)^2(n+2)} + \frac{2d}{3}}\right),$$

the following holds:

$$\left|U_{(i)} - \frac{i}{n+1}\right| + \left|U_{(i+1)} - \frac{i+1}{n+1}\right| < 2d.$$

Next, consider the difference between $U_{(i+1)}$ and $U_{(i)}$:

$$\left|U_{(i+1)} - U_{(i)} - \frac{1}{n+1}\right| \leq \left|U_{(i)} - \frac{i}{n+1}\right| + \left|U_{(i+1)} - \frac{i+1}{n+1}\right| < 2d.$$

Rearranging this inequality gives:

$$\left|U_{(i+1)} - U_{(i)}\right| - \frac{1}{n+1} \leq \left|U_{(i+1)} - U_{(i)} - \frac{1}{n+1}\right| < 2d.$$

Adding $\frac{1}{n+1}$ to both sides yields:

$$\left|U_{(i+1)} - U_{(i)}\right| < 2d + \frac{1}{n+1}.$$

By scaling back to the original variables, we obtain:

$$\left|X_{(i+1)} - X_{(i)}\right| < 2d(b-a) + \frac{b-a}{n+1}.$$

Next, we apply Hoeffding's inequality to the bounded random variable $U_{(i+1)} - U_{(i)}$, where $0 \leq U_{(i+1)} - U_{(i)} \leq b - a$ and we know that

$$\mathbb{E}\left[U_{(i+1)} - U_{(i)}\right] = \frac{1}{n+1}.$$

Using Hoeffding's inequality, we obtain:

$$\mathbb{P}(\left(\left|U_{(i+1)} - U_{(i)} - \frac{1}{n+1}\right| \geq d\right) < 2\exp\left(-\frac{2d^2}{(b-a)^2}\right).$$

Thus, with high probability, we have:

$$\left|U_{(i+1)} - U_{(i)}\right| < 2d + \frac{1}{n+1}.$$

Similarly, we find that:

$$\left|X_{(i+1)} - X_{(i)}\right| < 2d(b-a) + \frac{b-a}{n+1}.$$

Finally, we define δ as the minimum of the two bounds:

$$\delta = \min\left(2\exp\left(-\frac{d^2}{2\frac{(i+1)(n-i)}{(n+1)^2(n+2)} + \frac{2d}{3}}\right) + 2\exp\left(-\frac{d^2}{2\frac{i(n-i+1)}{(n+1)^2(n+2)} + \frac{2d}{3}}\right), 2\exp\left(-2d^2\right)\right).$$

Thus, with high probability, the difference between successive order statistics is bounded:

$$\left|X_{(i+1)} - X_{(i)}\right| < 2d(b-a) + \frac{b-a}{n+1}.$$

□

Lemma (Gap Bound for the Truncated Normal Distribution). Let $\sigma, b > 0$ and $X_1, \ldots, X_n \sim \mathrm{T}\mathcal{N}(0, \sigma^2, -b, b)$ be independent and identically distributed random variables. Consider an integer $m \in [n-1]$.

With probability at least $1 - \delta$, the following inequality holds:

$$|X_{(m+1)} - X_{(m)}| \leq 2d + |\mathbb{E}[X_{(m+1)} - X_{(m)}]|,$$

where

$$\delta = 2\exp\left(-\frac{2d^2}{(b-a)^2}\right),$$

and $\mathrm{Var}(X)$ denotes the variance of X.

Proof. To prove this lemma, it suffices to apply Hoeffding's inequality to the bounded random variable $X_{(m)}$. By considering the order statistics and the bounded range $[b, a]$, the gap between successive order statistics can be bounded as stated. The value of $|\mathbb{E}[X_{(m+1)} - X_{(m)}]|$ is calculated using numeric integration.

□